

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:23:12 ; Search time 46.26 Seconds
(without alignments)
4367.879 Million cell updates/sec

Title: US-09-762-311-5
Perfect score: 6055
Sequence: 1 MEPIFTARKHLIPNEVSD.....RSAEPSDRPEQTQPEPTGD 1168

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4976	82.2	1141	11	Q60949 mus musculus
2	3913	64.6	763	4	Q9UPP4 Q9UPP4 homo sapien
3	3463	57.2	674	4	Q96K82 Q96K82 homo sapien
4	2828	46.7	1299	4	O60343 O60343 homo sapien
5	2653	43.8	555	6	O97790 O97790 bos taurus
6	2404	39.7	494	11	Q923F8 Q923F8 mus musculus
7	1467.5	24.2	1379	5	Q9VNG9 Q9VNG9 drosophila
8	1462.5	24.2	1379	5	Q9TV11 Q9TV11 drosophila
9	1007.5	16.6	732	5	O24193 O24193 drosophila
10	580	9.6	997	4	Q9Y3P9 Q9Y3P9 homo sapien
11	523	8.6	1135	5	Q967F5 Q967F5 caenorhabdi
12	520	8.6	1140	5	Q967F5 Q967F5 caenorhabdi
13	510.5	8.4	607	4	O60546 O60546 homo sapien
14	510.5	8.4	810	4	O60447 O60447 homo sapien
15	500.5	8.3	794	4	O96CN4 O96CN4 homo sapien
16	492	8.1	1194	5	Q9VSI2 Q9VSI2 drosophila

17	479.5	7.9	803	5	Q9VYI9
18	465	7.7	894	3	Q12344
19	456	7.5	392	4	Q9H1Y9
20	429.5	7.1	809	11	P97366
21	421	7.0	876	3	Q9P4Y9
22	397	6.6	480	5	Q22729
23	386	6.4	356	10	O81347
24	385	6.4	304	10	Q9M894
25	367.5	6.1	749	4	Q96H01
26	367.5	6.1	777	4	Q95513
27	359	5.9	1005	5	Q9U2D8
28	358.5	5.9	511	5	Q9GPS2
29	338.5	5.6	868	4	Q9UPU7
30	333.5	5.5	828	4	Q92738
31	330.5	5.5	532	5	Q9Y1V8
32	328.5	5.4	533	4	Q9H822
33	326.5	5.4	537	11	Q91XR3
34	326.5	5.4	710	4	Q9H6A2
35	325.5	5.4	485	5	Q95RE0
36	325.5	5.4	571	5	O18357
37	325.5	5.4	908	5	Q23419
38	317	5.2	814	10	Q9M2W4
39	315.5	5.2	363	5	Q9VGL8
40	315	5.2	508	4	Q9BXI6
41	312	5.2	438	4	O76053
42	310	5.1	401	4	Q9NXX9
43	307.5	5.1	1291	5	Q9VPA6
44	306	5.1	614	5	O21928
45	303	5.0	804	5	Q9VFB6

ALIGNMENTS

RESULT 1

Q60949
ID Q60949 PRELIMINARY; PRT; 1141 AA.
AC Q60949; 1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TBC1.
GN TBC1D1 OR TBC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032578; PubMed=7566974;
RA Richardson P.M., Zon L.I.;
RT "Molecular cloning of a cDNA with a novel domain present in the tre-2
oncogene and the yeast cell cycle regulators BUB2 and cdc16.";
RL Oncogene 11:1139-1148(1995).
DR EMBL: U33005; AAA85223.1; .
DR MGD; MGI:1889508; Tbc1d1.
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR000050; PID_Domain.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00462; PTB; 2.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
DR PROSITE; PS01179; PID; 1.
SQ SEQUENCE 1141 AA; 129269 MW; 76AC52FCBB720816 CRC64;

Query Match 82.2%; Score 4976; DB 11; Length 1141;
Best Local Similarity 87.9%; Pred. No. 0;
Matches 970; Conservative 47; Mismatches 74; Indels 12; Gaps 5;
Qy 37 MPMLPWWVAEYRRLSRQSTKEPVTQVRLCVSPSLRCEPEPGRSQQWDFLYSSIFEC 96
||||| ||||||| :||| ||||| ||||||| :||| ||||| |||||

Db 1 MPMLPWVAEVRRLSGQCSKKEPRTKQVRLWVSPSGLRCEPDLEKSQPDPLICSIPEC 60
Qy 97 KPORVHKLHNSHDPYSFACLLKEDAVHROSQCYVFKADDTQKVPEIISIRQAGKIARQ 156
Db 61 KPORVHKLHNSHDPYSFACLLKEDAAHROSQCYVFKADDTQKVPEIISIRQAGKIARQ 120
Qy 157 EELHCPSEFDDTFSKFEVLFVFCGRVTVVAHKAPPALIDECIEKFNHVSGRGSESPRNP 216
Db 121 EELHCPSEFDDTFAKFEVLFVFCGRVTVVAHKAPPALIDECIEKFNHVSGRGRTD----- 174
Qy 217 PHAAPTGS-OEPVRRPMRKSFSQGLRSLAFRAKELQDGLRSGSFSSPEESDIENHLIS 275
Db 175 -WEAPTGPSAPGRPMRKSFSQGLRSLAFRAKELQDGLRSGSFSSPEESDIENHLIG 231
Qy 276 GNIVOPTDIENRMLFTIGSEVYLVISPDTKKTALEKNFEISFCOGIRHVDHFGPI 335
Db 232 GNIVOPTDIENRMLFTIGSEVYLVISPDTKKTALEKNFEISFCOGIRHVDHFGPI 291
Qy 336 CRESSGG--GGFHFVYVFCQNEALVDEIMMTLKAQFTVAAVQQTAKAPALCEGCPAQ 393
Db 292 CRECSGGSGGFHFVYVFCQNEALVDEIMMTLKAQFTVAAVQQTAKAPALCEGCPAQ 351
Qy 394 SLHKLCERTEGNNSKTKLELOKHLTTLTNOEQATIFEVQKLRPRNEORENELIISFLR 453
Db 352 GLHKLCERTEGNNSKTKLELOKHLTTLTNOEQATIFEVQKLRPRNEORENELIISFLR 411
Qy 454 CLYEEOKEHIIHGENKOTSQAAENIGSELPPSATFRFLDMLKNKAKRSLTESLESILS 513
Db 412 CLYEEOKEHSHGTAPKOTQVAEENIGSDLPSPASRFLDSLKNKAKRSLTESLESILS 471
Qy 514 RGNKARGLOHESISVDLSDSLSTLSNTSKPSVCEKEALPISSESKFLGSSDLSDDS 573
Db 472 RGNKARGLOHESISVDLSDSLSTLSNTSKELSMGDEKAPVSETSFKLGGSDDLSDS 531
Qy 574 ESHLPEEPAPLSQQAFFRRRANTLSHFPIECQEPPOPARGSPGVQKLMRYHSVSTETP, 633
Db 532 EGHIAEESALLSQQAFFRRRANTLSHFVECPAPPEAQSPGVSQKLMRYHSVSTETP 591
Qy 634 HERKDESKANHLGDSGPPVTKRHSWQOIFLRVATPQKACDSSSRVEDYSELGELPP 693
Db 592 HERKDESKANHLGDSGPPVTKRHSWQOIFLRVATPQKACDSSSRVEDYSELGELPP 651
Qy 694 RSPLEPVCEGDPGPPPEPKKRTSRELWQKAILQOILLRMKEKQKLAASENDLIN 753
Db 652 RSPLEPVCEGDPGQYKRRGRHASFESCGRRPSCSRSLYRMEKQKLAASENDLIN 711
Qy 754 KRLKLDYEITPCLKEVTTWVEKMLSTPORSKIKFDMKMKHSAVGGVPRHRHGEIWKPL 813
Db 712 KRLKLDYEITPCLKEVTTWVEKMLSTPGRSKIKFDMKMKHSAVGGVPRHRHGEIWKPL 771
Qy 814 AEQFHLKHOFPSKQOPKQVYKELLKOLTQSOQAILDLGRFTPTHPYFSAOLGAGQSL 873
Db 772 AEQFHLKHFPKQOPKQVYKELLKOLTQSOQAILDLGRFTPTHPYFSAOLGAGQSL 831
Qy 874 YNLIKAYSLLDQEVGYCOGLSVAGILLHMSSEEAFAKMLKFLMDGMLRKQYRPMIIL 933
Db 832 YNLIKAYSLLDQEVGYCOGLSVAGILLHMSSEEAFAKMLKFLMDGMLRKQYRPMIIL 891
Qy 934 QIOMYQLSRLLDYHRLDYNHLEEHIGPSLYAAPWFLTFWASQPLGFAVRFDMIFLQ 993
Db 892 QIOMYQLSRLLDYHRLDYNHLEEHETGPTTAAAPWFLTFWASQPLGFAVRFDMIFLQ 951
Qy 994 STEVIFKVALSLGSHKPLILOHENLETIVDTFKSTPLNGLVQMEKTIQVFMEDIAQ 1053
Db 952 STEVIFKVALSLGSHKPLILOHENLETIVDTFKSTPLNGLVQMEKTIQVFMEDIAQ 1011
Qy 1054 LQAEVEYHVLQELIDSDSPDNQMDKLEKTNLSLRKQNLDDLLEQOVANGRIQSLA 1113
Db 1012 LQAEVEYHVLQELIDSDSPDNQMDKLEKTNLSLRKQNLDDLLEQOVANGRIQSLA 1071
Qy 1114 TIEKLLSSSESKLKQAMLTLELER 1136
Db 1072 TVEKLLSSSESKLKQALTTLEVER, 1094

RESULT 2
Q9UPP4
ID Q9UPP4 PRELIMINARY; PRT: 763 AA.
AC Q9UPP4;
DC 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE KIAA1108 PROTEIN (FRAGMENT).
GN KIAA1108.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLIND-99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
DR EMBL: AB029031; BAA83060.1;
DR InterPro: IPR001220; Lectin_legB.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00164; TBC; 1.
DR PROSITE: PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 763 AA; 87585 MW; 32F7C3E2E3A49A83 CRC64;

Query Match 64.6%; Score 3913; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 2.4e-250;
Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 406 NSSKTKLEQLKHLTTLTNOEQATIFEVQKLRPRNEORENELIISFLRCLYEKQKEH 465
Db 1 NSSKTKLEQLKHLTTLTNOEQATIFEVQKLRPRNEORENELIISFLRCLYEKQKEH 60
Qy 466 IGEKMTQMAAENIGSELPPSATFRFLDMLKNKAKRSLTESILSRGNKARGLOEHS 525
Db 61 IGEKMTQMAAENIGSELPPSATFRFLDMLKNKAKRSLTESILSRGNKARGLOEHS 120
Qy 526 ISVDLSDSLSTLSNTSKPSVCEKEALPISSESKFLGSSDLSSESHPPEPAPUS 585
Db 121 ISVDLSDSLSTLSNTSKPSVCEKEALPISSESKFLGSSDLSSESHPPEPAPUS 180
Qy 586 PQQAFRRRANTLSHFPIECQEPPOPARGSPGVQKLMRYHSVSTETPHERKDFESKANH 645
Db 181 PQQAFRRRANTLSHFPIECQEPPOPARGSPGVQKLMRYHSVSTETPHERKDFESKANH 240
Qy 646 LGDSGTPVTKRHSWQOIFLRVATPQKACDSSSRVEDYSELGELPPSPLEPCEDP 705
Db 241 LGDSGTPVTKRHSWQOIFLRVATPQKACDSSSRVEDYSELGELPPSPLEPCEDP 300
Qy 706 FGPPPEEKRTSRELWQKAILQOILLRMKEKQKLAASENDLINRKLKLDYEITP 765
Db 301 FGPPPEEKRTSRELWQKAILQOILLRMKEKQKLAASENDLINRKLKLDYEITP 360
Qy 766 CLKEVTTWVEKMLSTPGRSKIKFDMKMKHSAVGGVPRHRHGEIWKFLAQHQPFS 825
Db 361 CLKEVTTWVEKMLSTPGRSKIKFDMKMKHSAVGGVPRHRHGEIWKFLAQHQPFS 420
Qy 826 KOQPKDVPYKELLKOLTQSOQAILDLGRFTPTHPYFSAOLGAGOLSYNLIKAYSLDQ 885
Db 421 KOQPKDVPYKELLKOLTQSOQAILDLGRFTPTHPYFSAOLGAGOLSYNLIKAYSLDQ 480
Qy 886 EVGYCOGLSVAGILLHMSSEEAFAKMLKFLMDGMLRKQYRPMIILQIQYQSLRLH 945

```

Db 481 EVGYCOGLSFVAGILLHLMSEEAFAKMLKFLMDFDGLRKQYRPMIILQIQMYQLSRLH 540
Qy 946 DYHDLNHLHEEHEIGFSLYAAPFLTMFAQFPLGFVARVDFMIFLQGTVEIFKVALS 1005
Db 541 DYHDLNHLHEEHEIGFSLYAAPFLTMFAQFPLGFVARVDFMIFLQGTVEIFKVALS 600
Qy 1006 LGSHPILQHLENLETIVDFTKSLPNLGLVQMEKTTINQFEMDIAKQAYEYHVHQ 1065
Db 601 LGSHPILQHLENLETIVDFTKSLPNLGLVQMEKTTINQFEMDIAKQAYEYHVHQ 660
Qy 1066 EELIDSPSLDNQRMKLEKTNSSLRKQNDLLQLQVANGRIQSLEATIEKLSSSKL 1125
Db 661 EELIDSPSLDNQRMKLEKTNSSLRKQNDLLQLQVANGRIQSLEATIEKLSSSKL 720
Qy 1126 KQAMLTLELERSALLQVVEELRRRSAPSDREPECTQPEPTGD 1168
Db 721 KQAMLTLELERSALLQVVEELRRRSAPSDREPECTQPEPTGD 763

RESULT 3
Q96K82 PRELIMINARY; PRT; 674 AA.
ID Q96K82
AC Q96K82;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE CNA FLJ1449 FIS, CLONE HEMB1001684.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBryo;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027355; BAB55057.1;
SQ SEQUENCE 674 AA; 77151 MW; 095584FB3C800A58 CRC64;

Query Match 57.2%; Score 3463; DB 4; Length 674;
Best Local Similarity 100.0%; Pred. No. 1.2e-220;
Matches 674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 495 MLNKKARSLTESLISRGNKARGLOEHSISVDLSLSSTLSNTPSKPSVCEKEALP 554
Db 1 MLNKKARSLTESLISRGNKARGLOEHSISVDLSLSSTLSNTPSKPSVCEKEALP 60
Qy 555 ISESSFKLLGSSDELSDSHSLPEEPAPLSPQQAFFRRANTLSHFPIEQEPPQAPARG 614
Db 61 ISESSFKLLGSSDELSDSHSLPEEPAPLSPQQAFFRRANTLSHFPIEQEPPQAPARG 120
Qy 615 PGVSQRKLMRYHSYSTPHERKDFESKANHLGDSGGTPVKTRHRSWRQIFLRVATPQK 674
Db 121 PGVSQRKLMRYHSYSTPHERKDFESKANHLGDSGGTPVKTRHRSWRQIFLRVATPQK 180
Qy 675 ACSSSSRYEDYSELGELPPRSPLEPVCEDEGPPGPPPEKRTSRELNOKAILQOILL 734
Db 181 ACSSSSRYEDYSELGELPPRSPLEPVCEDEGPPGPPPEKRTSRELNOKAILQOILL 240
Qy 735 LRMEKENKQLQASNDLNLNKLKLDYEITPCLKEVTVMKEMLSTPGRSKIKFDMKMH 794
Db 241 LRMEKENKQLQASNDLNLNKLKLDYEITPCLKEVTVMKEMLSTPGRSKIKFDMKMH 300
Qy 795 SAVGQGVPRHHRGGEIWKFLAQFHLKHQFSPKQKDPYKELLKQLTQSQAAILDLGR 854

```

```

Db 301 SAVGQGVPRHHRGGEIWKFLAQFHLKHQFSPKQKDPYKELLKQLTQSQAAILDLGR 360
Qy 855 TFPTHYPFSAQLGAGQSLYNILKAYSLLDQEVGYCOGLSFVAGILLHLMSEEAFAKMLK 914
Db 361 TFPTHYPFSAQLGAGQSLYNILKAYSLLDQEVGYCOGLSFVAGILLHLMSEEAFAKMLK 420
Qy 915 FLFMDGLRKQYRPMIILQIQMYQLSRLHLDYHRLDYNHLHEEHEIGFSLYAAPFLTMF 974
Db 421 FLFMDGLRKQYRPMIILQIQMYQLSRLHLDYHRLDYNHLHEEHEIGFSLYAAPFLTMF 480
Qy 975 ASQFPLGFVARVDFMIFLQGTVEIFKVALSLLGSHKPLLIQHLENLETIVDFIKSTLPNLG 1034
Db 481 ASQFPLGFVARVDFMIFLQGTVEIFKVALSLLGSHKPLLIQHLENLETIVDFIKSTLPNLG 540
Qy 1035 LVQMEKTTINQFEMDIAKQAYEYHVQLEELIDSPSLDNQRMKLEKTNSSLRKQ 1094
Db 541 LVQMEKTTINQFEMDIAKQAYEYHVQLEELIDSPSLDNQRMKLEKTNSSLRKQ 600
Qy 1095 LDLEQLQVANGRIQSLEATIEKLSSSKLKQAMLTLELERSALLQVVEELRRRSAP 1154
Db 601 LDLEQLQVANGRIQSLEATIEKLSSSKLKQAMLTLELERSALLQVVEELRRRSAP 660
Qy 1155 DREPECTQPEPTGD 1168
Db 661 DREPECTQPEPTGD 674

RESULT 4
O60343 PRELIMINARY; PRT; 1299 AA.
ID O60343
AC O60343;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KIAA0603 PROTEIN.
GN KIAA0603.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR EMBL: AB011175; BAA25529.1;
DR InterPro: IPR000050; PID_domain.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF006640; PID; 2.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00462; PTB; 2.
DR SMART: SM00164; TBC; 1.
DR PROSITE: PS01179; PID; 1.
SQ SEQUENCE 1299 AA; 146606 MW; 6EECC59BC62EEA88 CRC64;

Query Match 46.7%; Score 2828.5; DB 4; Length 1299;
Best Local Similarity 48.3%; Pred. No. 2.7e-178;
Matches 621; Conservative 184; Mismatches 291; Indels 191; Gaps 28;

Qy 21 FGLQLVGLSPVSHLTMPLPWWVAEVRRLSRQSTR---KEPVTQVRLCVSPSGLRCEP 77
Db 36 FRLWYGGSCLDHRTTLPLPWLMAEIRRRSQPEAGCGCAPAAREVILVLSAPFLRCVP 95
Qy 78 EPG-----RSQMDPLIYSIFECKPQRVHKLIHNSHDSYFACLTK--EDAVHRQ 126
Db 96 APGAGASGGTSPSATQNPVAF--IFEKHAHISRFTHNSHDLTYFAYLIKAQDPDQESQ 153

```



```
Db 424 VOMEKTSIOVFETDISKQOAYEYHVHVLQBELIDSSPLSDNQMDKLEKTNSSLRKNL 483
Qy 1096 DLLSQOVANGRIOSLEATIEKLSSSKLKQAMLTLELERSALLQTVELRRRSAPSD 1155
Db 484 DLLSQOVANGRIOSLEATVEKLTSBKLQATLALERSALLQTVELRRQTAELGS 543
Qy 1156 REPECTOPEPTGD 1168
Db 544 QESDPTLPKPSGD 556

RESULT 6
Q923F8 PRELIMINARY; PRT; 494 AA.
AC Q923F8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE UNKNOWN (PROTEIN FOR IMAGE:3500261) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004675; AAH04675.1;
FT NON-TER 1
SQ SEQUENCE 494 AA; 57041 MW; 4C1E7856F84A5F4F CRC64;

Query Match 39.7%; Score 2404; DB 11; Length 494;
Best Local Similarity 94.1%; Pred. No. 8e-151;
Matches 465; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

Qy 675 ACDSRSVEYDSELGEPSPLEPVCEDEGPPGPPPEKRTSRELRLQKATLQIILL 734
Db 1 ACDSPSRYEDYSELGEPSPLEPVCEDEGPPGPPPEKRTSRELRLQKATLQIILL 60
Qy 735 LRMEKENKQLQASENDLNRKLDYDEITPCLKEVTVVMEKMLSTPGRSKIKFDMKMH 794
Db 61 LRMEKENKQLQASENDLNRKLDYDEITPCLKEVTVVMEKMLSTPGRSKIKFDMKMH 120
Qy 795 SAVGQGVPRHRRGEIWKFLAEQPHLKHQFPKQKQDPVYKELKQTSQOHALIDLGR 854
Db 121 SAVGQGVPRHRRGEIWKFLAEQPHLKHQFPKQKQDPVYKELKQTSQOHALIDLGR 180
Qy 855 TFPHPHYSAQLGAGOLSLYNILKAYSILDOEVGYCGGLSFVAGILLHMSSEAFKMLK 914
Db 181 TFPHPHYSAQLGAGOLSLYNILKAYSILDOEVGYCGGLSFVAGILLHMSSEAFKMLK 240
Qy 915 FLFMDGLRKQYRPMIILQIQVLSRLLDHYHRLYNLHLEHEIGPSLYAAPFLTMF 974
Db 241 FLFMDGLRKQYRPMIILQIQVLSRLLDHYHRLYNLHLEHEIGPSLYAAPFLTMF 300
Qy 975 ASQPLGFLVAVFDMIFLQTEVIFKVALSLGSHKPLIIOHENLETIVDFIKSTPLNLG 1034
Db 301 ASQPLGFLVAVFDMIFLQTEVIFKVALSLGSHKPLIIOHENLETIVDFIKSTPLNLG 360
Qy 1035 LVOMEKTIQVFNENDIAKQOAYEYHVHVLQBELIDSSPLSDNQMDKLEKTNSSLRKN 1094
Db 361 LVOMEKTIQVFNENDIAKQOAYEYHVHVLQBELIDSSPLSDNQMDKLEKTNSSLRKN 420
Qy 1095 LDLLEQLQOVANGRIOSLEATIEKLSSSKLKQAMLTLELERSALLQTVELRRRSAPES 1154
Db 421 LDLLEQLQOVANGRIOSLEATVEKLTSBKLQATLALERSALLQTVELRRRSAPES 480
Qy 1155 DREPECTOPEPTGD 1168
Db 481 TPEPDCQTOLEPTGD 494

Query Match 24.2%; Score 1467.5; DB 5; Length 1379;
Best Local Similarity 32.1%; Pred. No. 2.9e-88;
Matches 372; Conservative 184; Mismatches 334; Indels 269; Gaps 32;

Qy 141 PEIISIRQAGK-----IARQEEHLCPESEFDDTFKSEFVLCGRVT 182
Db 67 PNPVSSKMKASKSYTHGLSSSGTGVNTPTSTSAQNSLLADISPNTHTFFEVVYVGKIR 126
```

```
RESULT 7
Q9VNG9 PRELIMINARY; PRT; 1379 AA.
AC Q9VNG9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PLX PROTEIN.
GN PLX OR CG1093.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE-20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003602; AAF51967.1;
DR FlyBase; FBgn004879; plx.
DR InterPro; IPR000050; PID_domain.
DR IntronPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00462; TBC; 1.
DR SMART; SM00164; TBC; 1.
SQ SEQUENCE 1379 AA; 154333 MW; 31DD6BE68CF4790E CRC64;
```

```
QY 183 VAHKAPALIDECIEKFN----- 201
Db 127 VSKRVPTFIDALPKFKAYDAQRLRLQNRKMSLSSEGGVIEAKPSSLSKSHDLKEE 186
QY 202 -----HVSGRSGSEPRPNPHAAPTGSOEVPVRRPMRKSF 236
Db 187 DEEEQOHKHHDSQDSQAKPLVOLQLTGAEEGAPRPLEDNKE---NKSPEKPLLRQ 243
QY 237 SQ-----PCLRSAPRKELO-----DGLRSSGFFSFEEESDIE-NH 272
Db 244 SGIELGHKSHSGSOPLAANSALFAPNVIVNQPTPRDQGVGTGTASAGPSQLHPNY 303
QY 273 LLSGHNIVOPTD---IEENRTMLTIGOSEVYLISPDTKKIALEKNKEISF 321
Db 304 AMD--NIPQRDSASQGIPIVVEQNRMTWFLVGRCDLRLSPDRKQVLLKDFKDVAS 361
QY 322 CSQGIHVDFHFGICRESSGGGFHFVCVYFOCTNEALVDEIMMTLKQAFVAAQQTAK 381
Db 362 CVHGKSLDHFGLICRELNDG---YIGVFKQSEHVCDIVAAIAQAFDTCA-EQKK 417
QY 382 APAQL--CEGCPLSLHKLCEIEGNSSKTKLEQKHQHTTTNOEQATIFEVOKLRPR 439
Db 418 QDTQIFSCHECHPMLWYHKLTQVGLSEKKQTALILRRIETLSDDQEITVWAKFCGSEKT 477
QY 440 NE--ORENELIISFRLCLYEKKHIGEMKOTSQMAAENIGSELPPSATRFLDMLK 497
Db 478 NSPVAEQNQLMLLRAHCESSQORHVH-----DTAENRSEFLNQYLGST-----IF 525
QY 498 NKAKRSLTESLSIRGKNGKAGLOEHSISVDLSLSTLNTSKEPVCVCEKALPISE 557
Db 526 MKAKRSLTNSFNLLKR-----KPS---KDDIAVPS 553
QY 558 SEFKLL--CSSEDLSDSESHLPEEPAPLSPOAARRRANTLSHPICEQPPQARGSP 615
Db 554 HNLRDIREGSAEPGLTQ-----SPPEGRSNTVTGASP-----SSKP 591
QY 616 GVSQRK--LMRVHSVSTETPHERKOPESKANHLGDSGGTPVKTRRHSWROQIFLRVATPQ 673
Db 592 TAEQLKSPMDIFIKVGNPK-----AETHQGSWQAILNSVVTPS 633
QY 674 KACDSSRYEDYSELGELPPRPVPCEDGPPGPPPEKKRTSRELRLWOKAILQQL 733
Db 634 KGLDSEVPTFELS-----PMRPAKRGKRDAAELRELWRTAIRQTIM 675
QY 734 LRMEKENKLOASENDLLKRLKLDYEITPCLKEVTVWEKML---STP-GRSKIKFD 789
Db 676 LRMETENAWLQARQENELKRIKLDYEIVPCDKOLIEREQIERNSTQIGNKK---D 732
QY 790 MEKMSAVQGVPRHRRGEIWFLEAQFHLKHQFP--SKQPK-DVPYKELLKQLTSQ 846
Db 733 PKVLGHAIPTGVPKRGDVTFLAQ-HSMNTAPVDTKRFFNTFYHMLLKLHTEHQH 791
QY 847 AILDLGRTPPHPTFSAGAGLSLYNLKAYSLDDQEVGYCQGLSPVAGILLHNS 906
Db 792 AIFIDLGRTPFHQFYKDPGLGQLSLFLNLLKAYSILDPGLGYCQGLGFCGVLLHCD 851
QY 907 EEAFLKFLMFDGMLRKQYRPMITLOIOMYQLSRLHLDYHRLYNHLEHEIGPSLYA 966
Db 852 ANSFQKLHLMFRNRMTKILFDMKKFQQLYQLSRLKVDHLPDLFVLDQNDVSTFLYA 911
QY 967 APWLTMFASQPLGPFVAVRFDMIFLQGTVEIFKVALSLGSHKPLILOHENLETIVDPI 1026
Db 912 APWLTVFSSQPLGPFVAVRFDLFLSSDVIFKFAIALLSVHKQQLAKDNFETIMDYL 971
QY 1027 KSTPLNLGLVQEKTNQVFMEDIAKQAYEVEHVHLOEEELIDSPSDNORMKLL--E 1084
Db 972 KTVVPMKMEHTCMEQINKLVFSDMIGKQLAEYKVEYNVLOEEI---TTNNHLEMLNRE 1026
QY 1085 KTNSSLRKONLDLLEOLOVANGRIQSLAETIEKLLSSSEKLAQMLTLEERSALLQTV 1144
Db 1027 KT-----QNHLEOQLQAQSSIAQLETT-----RSSQAQITTLQSQVOSLELTIQ 1073
```

```
QY 1145 ELRRSAEPSDREPECTOP 1163
Db 1074 TLGRYVQLVERHPDLELP 1092

RESULT 8
QYTY11 PRELIMINARY; PRT: 1379 AA.
AC QYTY11;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE POLLUX PROTEIN.
GN PLX OR POLLUX OR CG1093
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINAL;
RA MEDLINE=99030403; PubMed=9813038;
RA Xu X.Z.S., Wes P.D., Chen H., Li H.S., Yu M., Morgan S., Liu Y.,
RA Montell C.;
RT "Retinal targets for calmodulin include proteins implicated in
RT synaptic transmission."
RL J. Biol. Chem. 273:31297-31297(1998).
DR EMBL; Y17919; CAA76939.1; -.
DR FlyBase; FBgn0004879; plx.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC. 1.
DR SMART; SM00462; TBC; 1.
DR SMART; SM00164; TBC; 1.
SQ SEQUENCE 1379 AA; 154276 MW; F9F2074CDF014EBE CRC64;
```

Query Match 24.2%; Score 1462.5; DB 5; Length 1379;
Best Local Similarity 32.1%; Pred. No. 6.2e-88;
Matches 366; Conservative 174; Mismatches 322; Indels 277; Gaps 30;

```
QY 156 QELHCPSEFDDTFKKFVLCGRVTVAHKAPALIDECIEKFN----- 201
Db 100 QNLSILLADISPNNHTFFEVYVGVKIRVSKRPVNTFIDDLPKFKAYDAQRLRLQNRK 159
QY 202 -----HVSGRSGS 209
Db 160 MSLSEGGVIEAKPSSLSKSHDLKEEDEEEOHQHGHDDSDQSAKPLVOLQLTGAEEG 219
QY 210 ESPRPNPHPHAAPTGSQEPVRRPMRKSFQSLRSLAFRKELQDGGRLRSFGFSFEESDI 269
Db 220 AAPRPLEDNKE---NKSPEKPLLRQSQSIELG---HKEHSDGQSPSAA----- 262
QY 270 ENHLISGHNIV---OPTD----- 284
Db 263 -NSQLEAPNVIVNKQPTPRDQGVGTGTASAGPSQLHPNTAMDNIPQRDSASQGI 321
QY 285 ---IEENRTMLTIGOSEVYLISPDTKKIALEKNKEISFCQGIHVDFHFGICRESSG 341
Db 322 PYPVQNRMTWFLVGRCDLRLSPDRKQVLLKDFKDVASCVHGKSLDHFGLICRELNN 381
QY 342 GGGFHFCVYFOCTNEALVDEIMMTLKQAFVAAQQTAKAPAL--CEGCPLQLSLHKL 399
Db 382 DG---YIGVFKQSEHVCDIVAAIAQAFDTCA-EQKKQDTQIFSCHECHPMLWYHKLC 437
QY 400 ERIEGNSKTKLEQKHQHTTTNOEQATIFEVOKLRPRNE--ORENELIISFRLCLYE 457
Db 438 TDVEGLSEKKQTALILRRIETLSDDQEITVWAKFCGSEKTNPSVABEQNQLMMLRAHCE 497
QY 458 EKQEKHIHIGEMKOTSQMAAENIGSELPPSATRFLDMLKNAKSLTESLSILSRGNK 517
Db 498 SRQQRHVH-----DTAENRSEFLNQYLGST-----IFMKAKRSLTNSFNLLKR---- 542
```

```

QY 518 ARGLOEHSISVDLSSTLSNTSKERSVCEKALPISESSFKLL--GSSEDLSSDSSES 575
Db 543 -----KPS---KDDIAPVSHNLRDIREGSAEPLGTQ--- 570
QY 576 HLPPEPAPLSPOQAFRRANTLSHPPIEQEPQPARSGPGVSQRK--LMRYHSVSTETP 633
Db 571 -----SPEGRFSRNTVGSF-----SSKPTAEQLKSPMDIFIKVGNP 611
QY 634 HERKOFESKANHLGDSGGTPVTRHRSWROQIFLRVATPKQACDSSSYEDYSELGELPP 693
Db 612 KE-----AETHGSRWQAILNSVWTPSKGLDSEVPTEFLS----- 646
QY 694 RSPLEPVCEDPFGPPPEEKRTSRELRLAQKAILQIILLRMEKENQKLOASENDLLN 753
Db 647 -----PMRPAKRGKDAAEELRMWTAIRQIMLNRMETENAMLOARQENEL 695
QY 754 KRLKLDYEEITPCLEKVTVMKML---STP-GRSRIKFDMEKMSAVGQGVPRHHRGEI 809
Db 696 KRIKLDYEEIVPCDKQIERNQIERNSTQIGNKK---DPKVLGHAIRTVGPRSKRGDV 752
QY 810 WKFLAEQHLKHQFP--SKQPK-DVPYKELLKQLTSQOHAILEDLGRTPPHYFSAQL 866
Db 753 WTFLAEQ-HSMNTAPVDTKRPFNFTPYHMLLKLHTEHQHAIFIDLGRTPPHYFSAQL 811
QY 867 GAGQLSLYNILKAYSLLDOEVGYCOGLSFVAGILLHMESEEAFAKMLFELFMDGLRQY 926
Db 812 GUGQLSLNLLKAYSILDPGLGYCOGLFGICGVLLHLCDEANSFOLLKHLFRNRMTKY 871
QY 927 RPDMLILOTQYQLSRLHLDYHRLYNHLEHEIGPSLYAAPWFLTMFASQFPLGFVARV 986
Db 872 LPDMKKFQLQLYSLRYKDHLPDYVWLDQNDVSPTLYAAPWFLTMFASQFPLGFVARV 931
QY 987 FDMIFLOQTEVIFKVALSLLGSHKPLILOHENLETVDFIKSTLPNLGLVQMEKTINQVF 1046
Db 932 FDLLESSDVFIFKFAIALLSVHKQQLAKDNFEIMDYLTVPKMEHTCMEQIMKLVF 991
QY 1047 ENDIKOLQAYEVEVHVLQELIDSSPLSDNORMDKL--EKTNSSLRKQNLDLLEQLQVA 1104
Db 992 SMDIGKQLAEYVNEYVJQEEI-----TTNHLEMLNREKT-----QNHLEQLOLQA 1040
QY 1105 NGRIOSEATIEKLLSSSKLQAMLTLELERSALLOQVEELRRRSAPSDREPECTOP 1163
Db 1041 QSSIAQLETT-----RSSQQAQITTLQSQVQSLELTQTLGRYVQGVLEHNPDLPL 1092

RESULT 9
ID Q24193 PRELIMINARY; PRT; 732 AA.
AC Q24193;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE POLUX.
GN PLX OR CG1093.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OREGON R; TISSUE=TRACHEA, CNS;
RA Zhang S.D., Kassiss J., Olde B., Mellerick D.M., Odenwald W.F.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50542; AAB02200.1; -
DR FlyBase; FBgn0004879; plx.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS01179; PID; 1.
SQ SEQUENCE 732 AA; 83238 MW; 02707F4591775283 CRC64;

```

```

Query Match 16.6%; Score 1007.5; DB 5; Length 732;
Best Local Similarity 46.8%; Pred. No. 3.1e-58;
Matches 217; Conservative 79; Mismatches 137; Indels 31; Gaps 10;

QY 709 PPEEKRTSRELRLQKAILQIILLRMEKENQKLOASENDLLRKLRLDYEEITPCIK 768
Db 4 PAKRCKRDAAEELRMWTAIRQIMLNRMETENAMLOARQENELKRIKLDYEEIVPCDK 63
QY 769 EVTTVMKML---STP-GRSRIKFDMEKMSAVGQGVPRHHRGEIWKFLAEQHLKHQFP 824
Db 64 OLIEREQEIERNSTQIGNKK---DPKVLGHAIRTVGPRSKRGDVWTFLEAQ-HSMNTAP 119
QY 825 --SKQPK-DVPYKELLKQLTSQOHAILEDLGRTPPHYFSAQLGAGQLSLYNILKAYS 881
Db 120 VDKRPFNFTPYHMLLKLHTEHQHAIFIDLGRTPPHYFSAQLGAGQLSLYNILKAYS 179
QY 882 LLDQEVGYCOGLSFVAGILLHMESEEAFAKMLFELFMDGLRQYRPMIILQIQMYQLS 941
Db 180 ILDPGLGYCOGLFGICGVLLHLCDEANSFOLLKHLFRNRMTKYLPDMKKFQLQYQLS 239
QY 942 RLLHLDYHRLYNHLEHEIGPSLYAAPWFLTMFASQFPLGFVARVDFMIFLOGTEVIFKV 1001
Db 240 RLVKDHLPLDYVWLDQNDVSPTLYAAPWFLTMFASQFPLGFVARVDFLLESSDVFIEKF 299
QY 1002 ALSLLGSHKPLILOHENLETVDFIKSTLPNLGLVQMEKTINQVREMDIAKOLQAYEVEY 1061
Db 300 AIALLSVHKQQLAKDNFEIMDYLTVPKMEHTCMEQIMKLVFMDIGKQLAEYVNEY 359
QY 1062 HVLQBELIDSSPLSDNORMDKL--EKTNSSLRKQNLDLLEQLQVANGRIQSLAEATEIKLL 1119
Db 360 NVLQEEI-----TTNHLEMLNREKT-----QNHLEQLOLQFQSSIAQLETT----- 403
QY 1120 SSKLQAMLTLELERSALLOQVEELRRRSAPSDREPECTOP 1163
Db 404 --RSSQQAQITTLQSQVQSLELTQTLGRYVQGVLEHNPDLPL 445

RESULT 10
QY3P9
ID QY3P9 PRELIMINARY; PRT; 997 AA.
AC QY3P9;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE RAB6 GTPASE ACTIVATING PROTEIN, GAPCEN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99219856; PubMed=10202141;
RA Culf M.H., Possmayer F., Zander H., Bordes N., Jollivet F.,
RA Couedel-Courteille A., Janoueix-Lerosey I., Langsley G., Bornens M.,
RA Goud B.;
RT "Characterization of GAPCEN, a GTPase activating protein for Rab6,
RT part of which associates with the centrosome.";
RL EMBL J. 18:1772-1783(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Culf M.H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011679; CAB40267.2; -
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00640; PID; 1.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS01179; PID; 1.
SQ SEQUENCE 997 AA; 114142 MW; E3694B6FA9539C52 CRC64;

```

Query Match 9.6%; Score 580; DB 4; Length 997;
Best Local Similarity 23.9%; Pred. No. 9e-30;
Matches 221; Conservative 137; Mismatches 311; Indels 254; Gaps 33;

```
QY 295 IGSEVVLSPDKKIALEKNEKFEISFCGQIRHVDHFG-----FICRESSGGGFGHFV 348
DB 119 VSGIIVRLDPQNTTEIANPIYKILFCVNG-----HGTGPSDCFAFTESHYNABL-FR 172
QY 349 CYYFOCTNEALVDEIMMTLKQAFVAAVQOTAKAPQALCEGCPQLSHLKCERIEGMNS 408
DB 173 IHVERCEIQEAVSRILYSFATAPRSKATPLSATA--APQTPDSDFTFVSLE----- 225
QY 409 KTKLEQLQKHTTTNQBQATIFEVOKLPRNQRENNELIISLRC--LYEEKOKKHHI 466
DB 226 -----IKEDDGKGYFSAV-----PKDKRQ-----CFKLQGGDKKIVIV 260
QY 467 GEMKQTSMAAENIGSELPPSAFRFLDKLKNKAKRSLSLESILSRGNKARGLQEHSI 526
DB 261 VQQTNNKELAIERC-----FGL-----LLSPGKDVNSDMHLL 293
QY 527 SVLDSSLSLNTLSKPSVCKEALPISSESKLLGSEDSSESHEL-----PEE 580
DB 294 --DLES-----MGK-----SSDGKSVITGSWNPXS 317
QY 581 PAPLSQOAFRRRANTLSHPPIECQBPQPARGSPGVSRQKLMRYHSVSTETPHERKDFE 640
DB 318 P-----HFQV-----VNEETPKDKVLFM 335
QY 641 SKANHLGSGGTPVKTRHRSWRQOIFLRVATPQKACDSSSRVEDYSELGELPPRSPLEPV 700
DB 336 TTAVDL-----VIT-----EVQEPVRFLLTKVRVCSNPERL----- 367
QY 701 CEDGPGPPPEEKRRSRE-----LRELWKA-----ILQOILLRMKEKQKLO----- 745
DB 368 -----FWP-----FSKRSTTENFFLKQIKQKREKNTDTLYEVVCLSESESERRRKTAS 419
QY 746 -----ASNDLLNKLRLDYEE-----ITPCLKEVTVTWKMLSTPGRS 784
DB 420 PSVRLPQSGSQSVIISPDPDEEDENDEPILLSGSDGVSKCAEKILETWGELLS-KWHL 478
QY 785 KIKEDMEKMSAVGQVPRHHRGEIMKFLAEQFLKHQFPKQOQKDVPIYKELLKQTSQ 844
DB 479 NLNVRPKQLSSLYRNGVPEALRGVEVQWLLAGCHNDHLVEK-----YRIILITRESQP 530
QY 845 OHAILDLGRTFTPHYFSAQLGAGQLSYNLKAYSLLDQEVGYCGLSFVAGIILLHM 904
DB 531 DSAITRDINRTFPAHDYFKDTGGDQDSLYKICKAYSVDYEIGYCGQGSFLAAVLLHM 590
QY 905 SEEEAFKMLKFLMDCLMKQYRPMIILQIQMYQLSRLLHDYHRLDYNHLEHEIGPSL 964
DB 591 PESQAFSLVKIMFDYGLRFLKONFEDLHCKFYQLERLMOEYIPDLYNHFLDISLEAHM 650
QY 965 YAAFPWFLTFASOPPLGFFVARVDFMIFLAGTEYIFKVALSLGSHKPLIILQHENLETIVD 1024
DB 651 YASOWFLTLTAKEPLYMVPHIIDLICEGISVFNVALGLLAKTSKDDLL--TDFEGALK 709
QY 1025 FIKSTLPNLGLVOMEKTINGVFE-----MDIA-KOLQAYEYEVHVLQELI--DSSPLSDNQ 1078
DB 710 FFRVQLPK--RYRSEENAKKLMELACNMKISQKAKKYEYHTMREQQAQQEDPIERFE 767
QY 1079 RMD-KLEKTNSSLRKNLDLLEOVANGRIQSLIETIEKLLSSESKLKQAMLTLELERS 1137
DB 768 RERNRQOENNRLEQNDLDAHEL-----VTSKIALRKOLDNAEERKADALNKELMTKQ 821
QY 1138 ALLQTVTEELRRRSRSPDREPEC 1160
DB 822 KLIDAEERKRLSEESAKLEMC 844
```

RESULT 11
Q967F5

```
ID Q967F5 PRELIMINARY; PRT; 1136 AA.
AC Q967F5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F35H12.2B PROTEIN.
GN F35H12.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLIN=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Nelson J., Gattung S.;
RT "The sequence of C. elegans cosmid F35H12.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41540; AAK39232.1; -.
SQ SEQUENCE 1136 AA; 131257 MW; A86F4E3448103AE7 CRC64;
```

Query Match 8.6%; Score 523; DB 5; Length 1136;
Best Local Similarity 28.7%; Pred. No. 6.4e-26;
Matches 155; Conservative 80; Mismatches 181; Indels 124; Gaps 18;

```
QY 660 SNRQOIFLRVATPQKACD-----SSRYEDYSELGELPPRSP---LEPV 700
DB 347 AMQLEGVYFLTPPSKCDSDPNDRKLTFTISLSDSDRKRKQNLGKSPRMTQLLHPT 406
QY 701 -----CEDGPGPPPEEKRTSRELWQKAILQOILLRMKEKQKLOASENDLLNK 754
DB 407 GDDESDCDEPILLSGSKVSQCKEHEHMDQLI-----EN----- 442
QY 755 RLKLDYEEITPCLKEVTVTWKMLSTPGRSKIKFDMKMSAVGQVPRHHRGEIMKFLA 814
DB 443 -----WDQQSDRP-----QKISELYLDGIPDKLRGRVWOLLS 474
QY 815 EQPHLKHQFPKQOQKDV-PYKELLKQLTSQQHAILDLGRTPPTHYFSAQLGAGQLSL 873
DB 475 NAI-----DQDLVKEHYHFLSQPCPSEQIVMRDTHRTFPANDYKESQKQSOQL 525
QY 874 YNLIKAYSLLDQEVGYCGLSFVAGIILLHMSSEAFKMLKFLMDGMLRKQYRPMIIL 933
DB 526 YKISKVYSLVDEVSVCYQGLSFLAASLLHMPPEQAFCTLVKIMFNKGLRDLFKLGFDNL 585
QY 934 QIQMYQLSRLLHDYHRLDYNHLEHEIGPSLYAAPFLTFWFSQFPLGFFVARVDFMIFLQ 993
DB 586 HLRFFQTLTKDYIPDLSSHLEHIGIETHMYASQWFLTFTAKFFPLQWVFFFLDILFSLQ 645
QY 994 GTEVIEKVALSLGSHKPLIILQHENLETIVDFIKSTLPNLGLVQ-----MEKTINGVFE 1047
DB 646 GMTTIFHISALLDDAKTDLQLD-FEGTLKYFRVSLPRKYRTEASTKCLIHRAVK--FR 702
QY 1048 MDIAKOLQAYEYEVHVLQE-ELIDSSPLSDNQMDK-----LEKTNSSLRKQ-- 1093
DB 703 LNHSK-LEVYENEYKRIKELERENEDPV---LRMEKEIGHQANTLRERENDLDAHEL 758
QY 1094 --NLDLLEQLOVANGRIQSLIETIEKLLSSESKLKQAMLTLELERSAL--LQTVTEELRR 1149
DB 759 TSKIELRRRLDVAEDQIETSSANAERL-----TRQNMIDILEENKMLMEYEQIKEMYRR 812
```

```
RESULT 12
ID Q967F6 PRELIMINARY; PRT; 1140 AA.
AC Q967F6;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN F35H12.2B.
GN F35H12.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodirinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Nelson J., Gattung S.;
RT "The sequence of C. elegans cosmid F35H12.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41540; AAK39233.1;
SQ SEQUENCE 1140 AA; 131739 MW; E38D7163FAED5E99 CRC64;

Query Match 8.6%; Score 520; DB 5; Length 1140;
Best Local Similarity 28.7%; Pred. No. 1e-25;
Matches 155; Conservative 82; Mismatches 183; Indels 120; Gaps 18;

QY 660 SWRQOIFLRVATPQKACD-----SSSRVEDYSELGELPPRSP---LEPV 700
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 AWQLEGVYFLTPSKSCDQSDPNDRKLTFTISLESDSRKRKSKQNLGKSPRMPTQLLHPT 406
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 701 -----CEDGPFPPPEKRTSRELRLWKALQIQLLRMEKENQKLAQSENLLNK 754
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 407 GDESDCDEPLLSGSGKVSQCKEHELEMDQLI-----QKISELVLDGIPDKLRGRVWQLS 474
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 755 RLKLDYBEITPCLKEVTVVEKMLSTPGRSKIKFDMKMHSAVGGVPRHHRGIEWKFLA 814
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 -----WDQSDRP-----QKISELVLDGIPDKLRGRVWQLS 474
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 815 EOPHLKHQFSPKQPKDV-PYKELLKQLTSQOHAILEDLGRTPPTHPYFSAQAGAGQLSL 873
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 475 NV-----RILADQDLVEXYHIFLSPQSEQVIMDIHRTFFAHYFYESQKGQOQL 529
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 874 YNLIKAYSLDQEGYCOGLSFVAGIILLHMESEAPKMLKFLMDFMGLRKQYRPDMIL 933
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 530 YKISKVSLYDEEVSQOGLSFLAASLLHMPPEQAQCTLVKIMFNGLRDLFLKGLFDNL 589
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 934 QIQMYQSLRLHLDYHRLDYNHLESEHIGPSLYAAPWFLTMFASQPLGFVARVDMFLQ 993
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 590 HLRFFQTLALKDYIPDLSSHLEHIGIETHMYASQWFLTLTKAPLQWFFVLDLFLSQ 649
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 994 GTEVIFKVALSLGSHKPLLIQHENLETIVDFIKSTLPNLGLVQ-----MEKTINQVFE 1047
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 650 GMNTIFHISLALDDAKTDLQDLD-FEGTLKYFRVSLPRKYRTEASKYCLIHKAVK--FR 706
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1048 MDIAQLQAYEVEYHVLQE-ELIDSSPLSDNMDK-----LEKTNSSLRQ-- 1093
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 707 LNHSEK-LEVYENYKRIKELERENEDPV---LRMEKEIGRHOANTLRLERENDDLAHEL 762
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 1094 --NLDLLEQLQVANGRIQSLEATIEKLLSSSKUKQAWMLTLELSAL--LQTVVEELRRR 1149
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 763 TSKEIELRRKLDVADQIETSANATERL-----TRQNDLLEKNKMLREYEQIKEMYYR 816
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
ID O60546 PRELIMINARY; PRT; 607 AA.
AC O60546;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE TRUNCATED EVI5.
GN EVI5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Roberts T.P., Cowell J.C.;
RT "Truncation of the human EVI-5 gene within a neuroblastoma cell
RT line.";
RL Hum. Mol. Genet. 0:0-0(1998).
DR EMBL; AF042345; AAC16041.1;
DR InterPro; IPR000515; BPD_transp.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS00402; BPD_TRANS_P. INN MEMBR; UNKNOWN 1.
SQ SEQUENCE 607 AA; 70035 MW; FE18A5C190D8D21 CRC64;

Query Match 8.4%; Score 510.5; DB 4; Length 607;
Best Local Similarity 28.5%; Pred. No. 1.8e-25;
Matches 149; Conservative 99; Mismatches 171; Indels 103; Gaps 17;

QY 664 QIFLRVATPQKACDSSRYEDYSELGELPPRSPLEPCVEDGPGFPPPEEKRTSRELREL 723
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 44 QMASQVSPSTSLHTTSSSTLS---TPALSPSP-----SQLSPDVLEL 85
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 724 WQAKIQLIILLRMEKENQKLAQSENLLNKLKL-----DYEEITPCL-KEVT 771
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 LAKLEEQNILL---ETDSKSLRSVNGSRNNGSSSLVSSSSASSNLSHLEDSWILWGRIV 142
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 772 TVMEKMLSTPGRSKIKFDMKMHSAVGGVPRHHRGIEWKFLAEOFLHKLHQFSPKQPKD 831
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 143 NEMEDV-----RKKEKQVKEL---VHKGIPIHFAIVWQLCS---AQSMPIKQD--- 187
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 832 VPYKELLKQLTSQOHAILEDLGRTPPTHPYFSAQAGAGQLSLYNLIKAYSLLDQEGYCO 891
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 --YSELKMTSPCKLRIRDIARTYPEHNFPEKDKSLGQSEVLFNVKAYSLLVDREVGYCO 245
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 892 GLSFVAGIILLHMESEAPKMLKFLMDFMGLRKQYRPDMILQIOMYQSLRLHLDYHRLD 951
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 GSATFVGLLLMQPEEAPCFVFKLMQDYRLRELKFSMAELGLCMYQFECMIQHELP 305
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 952 YNHLEEHEIGPSLYAAPWFLTMFASQPLGFVARVDMFLQGTGTEVIFKVALSLGSHKP 1011
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 FVHFQSQSFHTSMYASSWFLTFLTPPLPVATRIDFIMSEGLEIVFRVGLALLQMNQA 365
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1012 LILQHENLETIVDFIKSTLPNLGLVQ-MEKTINQVFEINDIAKQLQ-AVEYEHVYLQELI 1069
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 ELMO-----LDMEGLMHQFKVPHQFGVDPDKLIQAAYQVY----- 403
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1070 DSSPLSDNQMDKLEKTNSSLRKONLLEQLQVANGRIQSLEATIEKLLSSSKLQKAM 1129
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 404 -----NSKMKKLEKTYTIKEME-EQVE-----IKRLTENLLAQRI 443
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1130 LTLELERSA-----LLQTVVEELRRRSAPSDREPECTOPE 1164
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 444 ETLEKHKCSSNYNEDFVLQLEKELVQ--ARLSEAESQCALKE 483
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
RESULT 14
O60447 PRELIMINARY; PRT; 810 AA.
AC O60447;
DT 01-AUG-1998 (TREMBlrel..07, Created)
DT 01-AUG-1998 (TREMBlrel..07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel..19, Last annotation update)
DE EVI-5 HOMOLOG
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN, IVER;
RA Roberts T.P., Cowell J.C.;
RT "Human EVI5 gene complete cDNA sequence.";
RL Hum. Mol. Genet. 0:0-0(1998).
DR EMBL; AF008915; AAC16031.1;-.
DR InterPro; IPR000515; BPD.transp.
DR InterPro; IPR000195; RabGAP.TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN.1.
SQ SEQUENCE 810 AA; 92847 MW; 8988011991329BF9 CRC64;

Query Match      8.4%; Score 510.5; DB 4; Length 810;
Best Local Similarity 28.5%; Pred. No. 2.6e-25;
Matches 149; Conservative 99; Mismatches 171; Indels 103; Gaps 17;

QY 664 QIFLRVATQKACDSSRYEDYSELGELPPRSPLEPVCEGDPGPPPEKKRTSRELRL 723
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 44 QMASQVASTSLHTTSSSTTSLTSSSTTSSSTTSSSTTSSSTTSSSTTSSSTTSS 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 724 WOKAIQIILLRMKEKNOKLOASENDLNLKRLK-----DYEEITPCL-KEYT 771
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 86 LAKLEQNILL---ETDSKLSVNGSRNRSGLSVSSSSASNSLSHLEEDSWILWGRIV 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 772 TVWEKMLSTPGRSKIKFDMKMHSAVGQVPRHHRGEIWKFLAEQFHLKHQFPKQPKD 831
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 143 NEMEDV-----BKKEKQVKEL--VHKGIPIHFRAIVWQLLCS-----AQSMPIKDQ--- 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 832 VPYKELLKQLTQSQAAILDLGRTPPTHPYFSAQLGAGOLSLYNILKAYSLLDQEVGYCQ 891
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 --YSELLKMTSPCEKLIIRDIARTIPEHNFKEKDSLGQEVLFNVMKAYSLVDREVGYCQ 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 892 GLSFVAGIILLHMSSEAFKMLFMDGLRKQYRPMIILQIQYLSRLLDHYHRLD 951
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 246 GSAFTVGLLLMQMPEEAFVFLMQDYRLRELKPKSMAELGLCMYQFECMIQEHLP 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 952 YNHLEHEIGPSLYAAPWFLTMFASQPLGFVARVDFMIFLOGTEVIFKVALSLGSHKP 1011
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 306 FVHFQSQSPHTSMYASSWFLTLFTTLPVPATRVDFIMFSEGIEIVFRVGLALLQMNQA 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1012 LILQHENLETIVDFIKSTLPNLGLVQ-MEKTINQVFEMDIKQLO-AYEVEYHVLQEELI 1069
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 366 ELMQ-----LDMGMLQHFQKVIPIPHQFDGVPDKLIQAYQVY----- 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1070 DSSPLSDNQMDKLEKNTSSLRKQNLDDLEQLQVANGRIQSLEATIEKLLSSSEK 1129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 404 -----NSKKMKLEKEYTIKTKEE--EQVE-----IKRLRTENRLLKQRI 443
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1130 LYLELERSA-----LIQTVEELRRRSAPSDREPTQPE 1164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 444 ETLKHKCSNTNEDFVLQLEKELVQ--ARLSEASQCALKE 483
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
Q96CN4 PRELIMINARY; PRT; 794 AA.
ID Q96CN4
```

```
Q96CN4;
AC 01-DEC-2001 (TREMBlrel..19, Created)
DT 01-DEC-2001 (TREMBlrel..19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel..19, Last annotation update)
DE SIMILAR TO ECOTROPIC VIRAL INTEGRATION SITE 5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA, AND CHORIOCARCINOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014111; AAH14111.1; -.
SQ SEQUENCE 794 AA; 91375 MW; 598B06DD2F098664 CRC64;

Query Match      8.3%; Score 500.5; DB 4; Length 794;
Best Local Similarity 26.4%; Pred. No. 1.2e-24;
Matches 146; Conservative 106; Mismatches 198; Indels 103; Gaps 16;

QY 669 VATPOKACDSSRYEDYSELGELPPRSPLEPVCEGDPGPPPEKKRTSRELRLWQKAI 728
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MASPTLSPDSSSQ-----EALSAPTC-----SPTSDSENLSPELE----- 36
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 729 LQIILLRMKEKNOKLOASENDL-----LNKRLKLDYEETPCLKE 769
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 37 -----LLAKLEQNRLLEADSKSRMNGSRNRSGLSVSSSSASNSLSHLEEDTWILWGR 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 770 TVTVMEKMLSTPGRSKIKFDMKMHSAVGQVPRHHRGEIWKFLAEQFHLKHQFPKQKP 829
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 93 IANEEWEN-----RRRKEKL-----LXELIRKGIPIHFRAIVWQLLCSATDM---PVKNQ- 139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 830 KDVPYKELLKQLTQSQAAILDLGRTPPTHPYFSAQLGAGOLSLYNILKAYSLLDQEVGY 889
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 ----YSELLKMTSPCEKLIIRDIARTIPEHNFKEKDSLGQEVLFNVMKAYSLVDREVGY 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 890 CGGLSFVAGIILLHMSSEAFKMLFMDGLRKQYRPMIILQIQYLSRLLDHYH 949
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 196 CGGSAFTVGLLLMQMPEEAFVFLMQDYRLRELKPKSMAELGLCMYQFECMIQEQ 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 950 DLYNLEHEIGPSLYAAPWFLTMFASQPLGFVARVDFMIFLOGTEVIFKVALSLGSH 1009
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 256 DLNTHFRSQSPHTSMYASSWFLTLFTTLPVPATRVDFIMFSEGIEIVFRVGLALLQVN 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1010 KPLIILQHENLETIVDFIKSTLPNLGLVQMEKTINQVFEMDI-AKQLOAVEYHVLQ--- 1065
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 316 QAEMLQDL-MEGMSQYFQRVIPHPQDSCDKLVKAYQVKNPKMKRLEKETEYAAKSK 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1066 -BELIDSSPL-SDN-----QRMDKLEK-----TNSSLRKQNLDDLEQLQVANGRIQSLEATIE 1116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 375 MEEQIEIKRLRTENRLLKQRIETLEKGVYTRAQAEAEENVYIKRELAVVYQCSSAAEDLQ 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1117 KLLSSEKLLQ-----AMLTLELERSALLQT-----VEELRRRSA 1151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 435 KAQSTIRQLQEQQENPRLTEDFVSHLETELEQSLRETETLGLALREMQDKVLDMEKRNSS 494
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1152 EPSDREPTQPE 1164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 495 LPDENNVAAQLQEE 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: August 28, 2002, 16:28:34
Job time: 322 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:23:52 ; Search time 15.15 seconds
(without alignments)
2985.111 Million cell updates/sec

Title: US-09-762-311-5
Perfect score: 6055
Sequence: 1 MPEITFTARKHLLPNEVSVD.....RSAEPSDREPTQPEPTGCD 1168

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	387	6.4	YD99_SCHPO	Q10496 schizosacch
2	343	5.7	YAD4_SCHPO	Q09830 schizosacch
3	317.5	5.2	YM52_YEAST	Q04322 saccharomyc
4	274	4.5	MIC1_YEAST	P53258 saccharomyc
5	247.5	4.1	YN33_YEAST	P48566 saccharomyc
6	232.5	3.8	CENF_HUMAN	P49454 homo sapien
7	227	3.7	MYHB_MOUSE	O08638 mus musculu
8	217	3.6	MYHB_HUMAN	P35749 homo sapien
9	212.5	3.5	MYHB_CHICK	P10587 gallus gall
10	208	3.4	GOG4_HUMAN	Q13439 homo sapien
11	207	3.4	CENE_HUMAN	Q02224 homo sapien
12	202	3.3	MYHB_RABIT	P35748 oryctolagus
13	197	3.3	MYHA_BOVIN	Q27991 bos taurus
14	196.5	3.2	TMF1_HUMAN	P82094 homo sapien
15	195	3.2	MYH9_HUMAN	P35579 homo sapien
16	195	3.2	MYHA_HUMAN	P35580 homo sapien
17	195	3.2	MYS3_SCHPO	O14157 schizosacch
18	193.5	3.2	MYSJ_DICDI	P54697 dictyosteli
19	193.5	3.2	MYH2_HUMAN	Q9ukp2 homo sapien
20	191.5	3.2	ACF7_HUMAN	Q9upn3 homo sapien
21	190.5	3.1	TPR_HUMAN	P12270 homo sapien
22	189	3.1	ALM1_SCHPO	Q9utk5 schizosacch
23	189	3.1	MYHA_RAT	Q9jlto rattus norv
24	187.5	3.1	YD86_SCHPO	Q10411 schizosacch
25	187.5	3.1	MYS2_DICDI	P08799 dictyosteli
26	186.5	3.1	HMW2_MYCGE	P47460 mycoplasma
27	186.5	3.1	MYH1_HUMAN	P12882 homo sapien
28	186.5	3.1	DMD_CANFA	Q97592 canis fami
29	186.5	3.1	AKA9_HUMAN	Q99996 h a-kinase
30	184.5	3.0	MYH9_RAT	Q62812 rattus norv
31	184	3.0	MYH7_PIG	P79293 sus scrofa
32	182.5	3.0	MYSN_DROME	Q98323 drosophila
33	182	3.0	MYH6_MESAU	P13539 mesocricetu

Query Match 6.4%; Score 387; DB 1; Length 1031;

Best Local Similarity 20.7%; Pred. No. 5.4e-13;

Matches 207; Conservative 153; Mismatches 377; Indels 262; Gaps 39;

ALIGNMENTS

RESULT 1

YD99_SCHPO

ID YD99_SCHPO STANDARD; PRT; 1031 AA.

AC Q10496;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Hypothetical 116.3 kDa protein C26F1.09 in chromosome I.

GN SPAC26F1.09.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: SOME, TO YEAST YMR192W AND YPL249C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z73100; CAA97366.1; -

DR InterPro; IPR000195; RabGAP_TBC.

DR Pfam; PF00566; TBC; 1.

DR SMART; SM00164; TBC; 1.

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 708 728 POTENTIAL.

FT CARBOHYD 113 113 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 155 155 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 355 355 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 438 438 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 449 449 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 1031 AA; 116289 MW; F2579A4C421D79EC CRC64;


```

QY 747 SENDLLNKRLLDYEEITPCLEKEVTTWVEKMLSTPGRSKIKFDMKMSHSAVQGVPRHRH 806
Db 458 LFFORLYKKYKIT-DEDTIGLIGTSSIGVK-----GRHG-KRMRHKKFRELYKNGVPLCYK 510
QY 807 GEIMKFLAEQFHLKHQFSPKQPKDVP---YKELLKQLTSQOHA-----ILIDGRFTFPH 860
Db 511 AKVWLECSGAQL-HS-----PGYBELLSRTDEVSASVAQIDMDINRTMAKNV 559
QY 861 YFSAQAGAGSLYNILKAYSLLDQEVGYCOGLSPV-AGIILLHMSSEAFKMLKFLMFD 919
Db 560 PFGGR-GPGIKRLRLVAYSRHNPHIGYCQGMNVIGAFLLLLYASEEDAFAVMLMSIEN 618
QY 920 MGLRKQYRPMILIOIQMYQLSRLLDYHRLYNHLEHEIGPSLYAAPWFLTFPASQFP 979
Db 619 VLPKRYFTPDLMTSRADQLVLSFKESLPEIYSHLELGLVDLDAISFHWFLSVYTDLP 678
QY 980 LCFVARVDFMFLQGTVEIFKVALSLGSHKPLILIOHENLETVDFIKSTLPNLGLVQME 1039
Db 679 TNISRIEDMLFCDGYVCLFRVALTILKSLKQIILACNSSAIVSFLSDL----- 728
QY 1040 KTIQVNFEMDIKQVAYEYHVLQELIDSSPLSDNORMDKLEKTNSSLRKQNL 1095
Db 729 -----VOYSFQDPSFIKEAA-----DRWSKLVTEKSIERKRN 761

RESULT 3
YMS2_YEAST STANDARD PRT; 720 AA.
AC Q04322;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 82.1 kDa protein in SGS1-MRPL24 intergenic region.
GN YMR192W OR YMR5646.04.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YPL249C AND S.POMBE SPAC26F1.09.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 247815; CAA87813.1;
DR SGD; S0004804; YMR192W
DR InterPro; IPR000195; RABGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
DR Hypothetical protein.
SQ SEQUENCE 720 AA; 82131 MW; E1423DB4F15F7267 CRC64;

```

Query Match 5.28; Score 317.5; DB 1; Length 720;
 Best Local Similarity 21.08; Pred. No. 1.5e-09;
 Matches 149; Conservative 126; Mismatches 244; Indels 191; Gaps 26;

```

QY 557 ESSFKLLGSSDLS--DSSEHLPE-----PAPLSPPQAFRRRANTL--SHFPICQE 606
Db 92 ESQKTIPLSDLSQLETESKVEEALKRITSPPLP-----RADCIESAALKSSL 145
QY 607 PP-----OPAGSPGVSGVSKLMRYHSVSTETPHERKDFESKANHLGDSGGTPVKTRR 658
Db 146 PPVLAGNKNDQAPLDRPOLPRQVNAETLHKAPHGN----- 183

```

```

QY 659 HSWROQIFLRVATPOKACDSSSRVEDYSELGELPPR-----SPLEPVCEDEGPGPPPEKKR 715
Db 184 -----ATPSKSPTSVAGNSSSTPTPLPRRIEDPLDLAAQ----- 219
QY 716 TSRELRELWQKAILQOILLRMEXENOKLOASENDLLNKRLLKLDYEEI--TPCLKEVTV 773
Db 220 -----KHFLASTFRKRMFLFYKSEDSIKCDLDKN-ILN--LKDSKKINNNEIPEVSSF 271
QY 774 WEKMLSTPGRSKIKFDMKMSHSAVQGVPRHRGIGWFLAEQFHLKHQFSPKQPKDVP 833
Db 272 WLKLVIGDYQNTLIN-DIETLHFLQLSRGIPAAAYRLVWQVLSV-----AKSFSFPI 321
QY 834 YKELLKQLTSQOHAILEDLGRFTFPHYFSAQLGAGQLSLY-----NLTKAYS 881
Db 322 YETVLTMAP-----FDVQEFENQLKMDMEVPSEYVVKRISNVLKAYL 363
QY 882 LLDQEVGYCOGLSFVAGIILLHMSSEE--AFKMLKFLMFDGLRKQYRPMILIOIQMY 939
Db 364 LFDPECEFTDIAIYIN-MILDVCEEAEANAFGLVRLMKVYGLRLLFLPSASEIDILCYK 422
QY 940 LSRLLHDYHRLYNHLEHEIGPSLYAAPWFLTFASQFPLGFVARVDFMFLQGTVEIF 999
Db 423 FDLRVEEYPEIHNHMYEKGYRSSMFLPGFPTTLFQKKLPTFQIOPRIGDMVFLGIDSIM 482
QY 1000 KVALSLGSHKPLILIOHENLETVDFIKSTLPNLGLVQMEKT-----INQVF 1046
Db 483 RILATLLSNSRDLHK-MGFDDMLELLKSLGLLDAYIKQNDCTRGDTLLSNCEMDKLQDS 541
QY 1047 EMDIA---KQLOAYEVEYHVL-----QEELIDSSPLSDNORMDKLEKTNSSLRKON 1094
Db 542 MMKVAITPKTKMKYSSEYEEIHRLDNKEVQYKSTENKHLQKHVRKLENDYTSLSREH 601
QY 1095 LDLLEQ-----LQVANG-RIQSLEATIEKLSSSEK-----LKQ 1127
Db 602 VTIANELVKRLNTIESVLNENNGYKLIQILD--LKKLDSEKKQVLGVVVPNDLKKOLEE 659
QY 1128 AM-----LTLELERSALLQTVELRRRSAPSDREPECTQPEPTG 1167
Db 660 TMKKNTQVMDENLKLQDRISLERLIEIKTANKNGTLFEYSNKNPNPLG 709

```

RESULT 4

```

MICL_YEAST STANDARD PRT; 950 AA.
AC P53258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MIC1 protein.
GN MIC1 OR YGR100W OR G5717.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Serpe M., Kosman D.J.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RA Hernandez K., Weber N., Wipfli P., Schmidheini T.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERACTS WITH MAC1.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO S.POMBE SPAC48.04, SPCC1259.11C, SPBC215.01 AND
CC YEAST YNL293W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U52667; AAB01977.1; -
DR EMBL: 272885; CAA97103.1; -
DR SGD: S0003332; MDR1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR004182; GRAM.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF00036; ehand; 1.
DR Pfam: PF02893; GRAM; 1.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00164; TBC; 1.
KW Transmembrane.
FT TRANSMEM 121 141 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
SQ SEQUENCE 950 AA; 109259 MW; 888BC930C1358657 CRC64;

Query Match 4.5%; Score 274; DB 1; Length 950;
Best Local Similarity 23.3%; Pred. No. 4e-07;
Matches 98; Conservative 73; Mismatches 162; Indels 88; Gaps 12;

QY 714 KRTSRELRLWOKA-----ILQIILLRMEKE-----NOK-----743
DB 110 KRVERSPSEYEFALLVLTGAKVLQIGIRYSEQFCDKLNLKENIPNAKTLPAF 169
QY 744 LQASENDLLNKRUKLDYEIT-----PCL-----KEVTYVWEKMLSTPGRSK 785
DB 170 LETSYSEFLIAKNILGKDTITPRAGLQGHFKYKPGNPTVMVKEKAKRLWFDYFRENGRNL 229
QY 786 IKFDMKMHSAVCGQYPRHRRGEIWEFLAQFHLKHQFSPKQOPKDPVYKELLQK---LT 842
DB 230 AVVQTPMFKRLIRIGYPRMRGRGIEWELSCGMYMRYANSGE-----YERILNENAGT 282
QY 843 SQO-HAILIDLGRFTTHYFSAQAGOLSLYNILKAYSLLDQEVGYCQGLSFVAGILL 901
DB 283 SQAIDEERDKLSPEYSAQTGEGIQLR--NVLTAYSWKNPDVGYCQAMNIVVAGFL 340
QY 902 LHMSEEAFAKMLFMDFMGLRKQYRPMITIIQIOMYQLSRLLHDYHRLYNHLEHETG 961
DB 341 IFMSEQAFCWCLNLC-DIIVPGYYSKTYGTLDDQRFESFEDRMPVLWEYILQHDIQ 399
QY 962 PSLYAAPWELTASQFPLGFAVFDMLFLOCTEVEIFKVALSLGSHKPLIIQHENLET 1021
DB 400 LSVSLPWFSLFSTSMPLLEYAIRMIDIFFMNGSITLFOVALAVLKINADDILQADDDGM 459
QY 1022 IVDFKSTLPNLG-----LVQMEKTIINQVFEMDIKQLOQAYEVEYHVLQEE 1067
DB 460 FIAIKHYFTLQSAHPDSSDIKYRIKQELLYTAEK-----EFSVISEE 507
QY 1068 L 1068
DB 508 M 508

RESULT 5
YN33_YEAST
ID YN33_YEAST STANDARD; PRT; 633 AA.
AC P48566;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 73.0 kDa protein in CLAA-PUS4 intergenic region.
GN YNL293W OR N0470.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=S288C / FY1679;
RX MEDLINE=96132033; PubMed=8553702;
RA Maurer K.C.T.; Urbanus J.H.M.; Planta R.J.;
RT "Sequence analysis of a 30 kb DNA segment from yeast chromosome XIV
RT carrying a ribosomal protein gene cluster, the genes encoding a
RT plasma membrane protein and a subunit of replication factor C, and a
RT novel putative serine/threonine protein kinase gene.";
RL Yeast 11:1303-1310(1995).
CC 1- SIMILARITY: TO YEAST MIC1 AND S.POMBE SPAC4G8.04.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U23084; AAC49106.1; -
DR EMBL: 271569; CAA96211.1; -
DR SGD: S0005237; YNL293W.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00164; TBC; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 422 442 POTENTIAL.
SQ SEQUENCE 633 AA; 72999 MW; 5F62607B95BBAFB9 CRC64;

Query Match 4.1%; Score 247.5; DB 1; Length 633;
Best Local Similarity 23.2%; Pred. No. 5.7e-06;
Matches 78; Conservative 72; Mismatches 91; Indels 95; Gaps 15;

QY 738 EKENOKLQASENDLLNKRUKLDYEITPCLKEVTYVWEKMLSTPG-----RSKIKFDM 790
DB 161 KQKNYISEAYD---KMWVEYSQY--CVRR-KHKWQLLEKSLGVLPTDDSPSRFSPKS 213
QY 791 EKMHSVAGVGVRHRRGEIWE-----KFLAEQFHLKHQFSPKQOPKDV 832
DB 214 ERKRYVRKGIPEWNGNAWHFARGQEKLNKNKGYSOLLKMKQIKQKNQNEKQVQLD 273
QY 833 PYKELLKQLTSQOHAIDLGRFTTHYFSAQAGAGQ-----LSLYNLKAYSLLDQEV 887
DB 274 ---DIIER-----DLNRTFPDNIHFQSSLHNKEGPPTIKSLRRVLVAFSLNPKI 320
QY 888 GYCQGLSFVAGIILLHMSSEEAFAKMLKFLM-----FDMGLR-----KQYR 927
DB 321 GYCQSMNFTAGLLFLDEERAFWMLVIITSRYLPGVHNINLEGVNIQDGLVLMCVKEYI 380
QY 928 PDMIIQIOMYQLSRLLHDYHRD-----LYNHLEEHE-----IGP-SLYAAPWF 970
DB 381 P-----EWSYIKPSIDHOKNNKTFSPSNKVKLFN-MQKNEFLYRLPITLCTASWF 432
QY 971 LTMFASQFPLGFAVFDMLFLOCTEVEIFKVALSLL 1006
DB 433 MSCFVGVPDIETTLRIWDCLFYEESHFLPKVSLAVL 468

RESULT 6
CNF_HUMAN
ID CNF_HUMAN STANDARD; PRT; 3210 AA.
AC P49454; Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CNF-F kinetochore protein (Centromere protein F) (Mitosisin) (AH
DE antigen).
GN CNPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A.
 RC TISSUE-Breast carcinoma;
 RA MEDLINE=953348175; PubMed=7542657;
 RX Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
 RT "CENP-F is a protein of the nuclear matrix that assembles onto
 RT kinetochores at late G2 and is rapidly degraded after mitosis.";
 RL J. Cell Biol. 130:507-518(1995).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=95379848; PubMed=7651420;
 RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
 RA Jones D., Yang-Feng T.L., Lee W.-H.;
 RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
 RT that is specifically involved in mitotic-phase progression.";
 RL Mol. Cell. Biol. 15:5017-5029(1995).
 [3]
 RN SEQUENCE OF 2194-3210 FROM N.A.
 RP MEDLINE=95336446; PubMed=7612011;
 RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
 RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
 RT domain sufficient for nuclear localization.";
 RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
 [4]
 RN CHARACTERIZATION.
 RP MEDLINE=95370296; PubMed=7642639;
 RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
 RT "The C terminus of mitotin is essential for its nuclear localization,
 RT centromere/kinetochore targeting, and dimerization.";
 RL J. Biol. Chem. 270:19545-19550(1995).
 [5]
 RN CHARACTERIZATION.
 RP MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC -I- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN
 CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
 CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
 CC -I- SUBUNIT: HOMO- OR HETERODIMER.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),
 CC REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF
 CC THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
 CC -I- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
 CC -I- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: U19769; AAA82889.1;
 DR EMBL: U30872; AAA82935.1;
 DR EMBL: U25725; AAA86889.1;
 DR HSP: P02649; 1LE4.
 DR MIM: 600236;
 KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
 KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
 FT DOMAIN 14 197
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 273 769
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 823 1328
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1642 1746
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1862 2987
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 2207 2568
 FT REPEAT 2207 2386
 FT REPEAT 2389 2568
 FT DOMAIN 3015 3032
 FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT T -> A (IN REF. 2).
 FT L -> Q (IN REF. 2).
 FT CONFLICT 16 250

FT CONFLICT 272 272
 FT CONFLICT 611 611
 FT CONFLICT 1494 1589
 FT CONFLICT 1611 1611
 FT CONFLICT 1811 1811
 FT CONFLICT 2242 2243
 FT CONFLICT 2335 2335
 FT CONFLICT 2492 2492
 FT CONFLICT 2545 2561
 FT CONFLICT 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
 SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
 Query Match 3.8%; Score 232.5; DB 1; Length 3210;
 Best Local Similarity 19.3%; Pred. No. 0.0003;
 Matches 240; Conservative 191; Mismatches 440; Indels 373; Gaps 55;
 QY 74 RCEPEGRSQWDPLIYSIFECK-PQRVHKLIHNSHPS-YFACLIHEDAVHROSICVY 131
 DB 120 RCKSELRSSQAAQADSVLNPCTPQ---KIFTPLTPSQIYSKYEDLKEK-----YN 172
 QY 132 FKADDTKVPEIISIRQAGKIARQELHCPSEFDDTFKKEFVLCGRVTVAAHKAPPA 191
 DB 173 KEVEERKLEAEVKALQ-----AKK-----ASQTLPOA 200
 QY 192 LIDECIEKFNHVGSRGSESPRNPHPAAPTGSOPVRRPMKKSQPGCLSLAPRKELQ 251
 DB 201 TMNHRDIARHQASSVFSWQOEKTPSHLSSNSQRTPIRRDFSASYFSGELEVTSPRSTIQ 260
 QY 252 DGGL-RSSGFFSS-----FEESDIENHL-----ISGHNIVQPTDIEENRTMLF 293
 DB 261 IGRDANSFFGNSSPHLLDQLKAQNOELRANKINLELRLOQGHKEKMGKQVKNKQELQL 320
 QY 294 TIGOSEVLIISPD-----TKKIALEKNFKEISFCSGIRHVDHF 332
 DB 321 QLEKAKVELIEKEKVLNKRDELYRTTAQYDOASTKYTALEQKLAKT-----E 369
 QY 333 GFICRESSGGGFHVCVVFQCTNEALVDE-----IMWILKQAFVAAVQO 378
 DB 370 DLSCORQNAESA-----RCSLFQIKKEKEFEELSROORSFOTLDO-----ECOM 417
 QY 379 TAKAPALCEGCPLOSLKICE-RIEGMNSSKTKLE--LOKHLLTLTNOBOATIEEYQK 435
 DB 418 KARLTQELQQA---KNMNVLOAELDKLTSVQKQLENLEEFKQKLCRAEQAFQASQIKE 474
 QY 436 LRPRNEORENELIISFLRCLYEKKQKEHIHI-GENKOTSQM--AAENTGSELPPSATFR 492
 DB 475 NELRRSMEEMKENLLKSHGEQKAREVCHLEAEKLNKQCLNQSQNAEEMKAKNTS-Q 533
 QY 493 LDMLK-----NKAERSLT-ESLESILSRGNKARCL-----QHSISVDLSSLSSTL 538
 DB 534 ETMLRDLOEKINOENSLTLEKLLAVADLEKQDCSDLLKKREHHI-----EQLNDKL 588
 QY 539 SNTKEPSVCEKEALPISESSFKLIGSSSEDLSSSESHLPEEPAPLSPQQAFFRRANTLS 598
 DB 589 SKTEKES-----KALLSALELKKYEELKEEKTLS----- 620
 QY 599 HFPIECQPPQAPAGSPGVQSKMLMYHSVSTPHEPKDFESKANHLGDSGGTPVKTRR 658
 DB 621 -----CWKS-----ENEKLL-----TQMESEKENLQSKINHL-----ETCLKTQ 655
 QY 659 ---HSW-----RQOIFLRVATPOKACDSSSRDYSELGELPPRSPLEPVCEGDP 705
 DB 656 IKSHEYNERVTEMLDMRENLSVEIRNLHNVLDSSVETQKLAYMEIQQKAE----- 708
 QY 706 FGPPPEKKRTSRELRELWQKAIQQIILLRMKEKNQKLAQSENDLLNK-----RLKIDY 760
 DB 709 FSDQKHQK-----ELENMCLKTSQLTQGVDELEHLQLLSNEINDKRCYQDLHAHY 760
 QY 761 EEITPCPKKE-----VTTVWEKMLSTPGRSKIKFDME-KMH-----SAYG-QGVPRHREI 809
 DB 761 ESLRDLKSKDASLVN-----EDHQSLLAFDQOPAMHHSFANIIGEQSMPSEK 814

Qy 243 SLAFKELQDGLSSGFFSFEES-----DIENHLISGH-----NIVQPTDI 285
Db 962 EEAARQKLEKVTAAEKIKLEDDILVMDQNSKLSKRLLEBVSQDLTTNLAEEEEK 1021
Qy 286 EENRMLFT-----IGQSEVILSPDTKKIALENKREKREISFCSSQGIHVDFHGFICRESS 340
Db 1022 AKNTTKLKSKEHSMISELEVLKKEEKSQRELEKLKRL-----EG 1062
Qy 341 GGGGFHFVYVQCTNEALVDEIMMTLKQAFVAAVQOTAKAPQLCEGCPQLSLHKLCE 400
Db 1063 DASDFH-----EQIAD-----LQAQIAELKQWLAKKEEL-----QAALARLDE 1101
Qy 401 RIEGNSKTKL-ELQKHLTL-----TNOEQATIFEVQKLRPRNE-----441
Db 1102 EIAQNNALKKIRELEGHISDQLQEDLDSERAARKAOKRDLGCELEALKTELEDTLDS 1161
Qy 442 -----QRENELISFLCLYEKKEHIGEMKQTSQMAAENIGSELPP-SATRF 491
Db 1162 TATQOELRAKREQVTV-LKKALDEETRSHEAQVOEMKQKHTQAVEELTEQLEQFKRA 1220
Qy 492 RLDMLK-----NKAKRSL-----TESLESILSRGNKARGL---521
Db 1221 NLDSKQTFLEKNADLACELVGLQAKQEVHEKHKKLEVLQDLQSKCSDGERARAEUSD 1280
Qy 522 QEHSISVDLDS-----SLSTLSNTSKEPSVCKEALPISSESKF 561
Db 1281 KVHKLQNEVESVTGMLREAEKATKAKDVASLQSLQDQTLQELQETROKLNVSFK---1337
Qy 562 LIGSEDLSSSESHLPEAPLPQQAFFRANTLSHFPIECQFPPOPARSGVGSORK 621
Db 1338 -LRQLEDRNSLQDLDEE---MEAKONLHRVSTLN---IQLSD-----SKKK 1379
Qy 622 LMRHVSSTETPHE-----RKDFESKANHLGDSGTPVKTRRHSWR-----662
Db 1380 LQDPAS-TIEWEBGKKRLQEMEGLSQOYEEKAAAYDKLEKTKNRQLQELDLVVDLDN 1438
Qy 663 -QQFLRVATOKACDS-----SSRYEDYSELGELPPRSPLEPCVEDGDFPGPPPEEK 713
Db 1439 QRLVSNLEKQKQKFDOLLAEKNISSKYAD-----ER 1471
Qy 714 KRTSRELQKAI-----LQOILLR--MEKENOKLOASENDLLNKR--LKLDYEIT 764
Db 1472 DRAEAREKTKALSARALEEALKEELERTNKLKAMEDLVSSKDDVGRNVHLE 1531
Qy 765 PCLKEVTVWEKMLSTPGRSIKFKDMEKMSAVGGVPRHRGETWFKPLAEQFLKHKQFP 824
Db 1532 KSKRALETQMEEM-----KTQLESEDDVQATEDAKLRLEV-----NQALKGQFE 1577
Qy 825 SKQPKQVPYKELKQLTSQOHALIDLGRTFPHYPSAQLGAGQLSYLNILKAYSLLD 884
Db 1578 RDLQARDQNEKRRQLOROLHEYETEL-----EDERKQRAALAAAKKLEGLDKLDEL--1631
Qy 885 QEVGVCGLSFPVAGILLHMSSEEAFAKMLFLPDMGLRKQYRPMITLQIQYLSLL 944
Db 1632 -----QADSIAIK-----REAIKO-----LRK-----LQAKMDFOREL 1661
Qy 945 HDYH--RD--LYNHLHEIGPSLYAAPWFLTMFASQPLGFARVDFMIFLOQTEVIFK 1000
Db 1662 DDARASDEIFATSKENKAKSLEA-----DLMLQLEDLAAERARKQADEKEELAE 1716
Qy 1001 VALSLGSHKPLIIOHE--NLETIVDFTKTLPNLGLVQMEKTNQVFMIDIAKOLQAYE 1058
Db 1717 LASSLSGRN--TLQDEKRRLEA-----RIAQLEEELE---EEQGNMEANSR 1758
Qy 1059 VEYHVLOEELSDSPSLDNQMDKLEKTNSSLRKNQDLLEQLQVANGRIQ-----S 1110
Db 1759 VRKATLQAEQLSNELATERSTAQNESARQQLERKNELSKLQEVGEVAKAKLSTVAA 1818
Qy 1111 LEATL-----EKLSSSKLQAMLTLLERSALQTVLEELR-----1147
Db 1819 LEAKIAQLEQVEQAEKQKATKSLKQKDKLKEVLQVEDERKMAEQYKQAEKGNTK 1878

Qy 1148 -----RRSAEPSDREPEC 1160
Db 1879 VKQLKQLEAEESQ 1895
RESULT 8
MYHB_HUMAN
ID MYHB_HUMAN STANDARD: PRT: 1972 AA.
AC P35749: O00396; P78422; O94944;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMHC).
GN MYH11 OR KIAA0866.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
Fuhmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
Raichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
human chromosome 16p and 16q";
RL Genomics 60:295-308(1999).
RN (2)
RP SEQUENCE OF 1-1266 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro";
RL DNA Res. 5:355-364(1998).
RN (3)
RP SEQUENCE OF 885-1972 FROM N.A.
RX MEDLINE=93263189; PubMed=7684189;
RA Matsuo K., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,
Yanagisawa M., Masaki T., Takao A.;
RT "Human smooth muscle myosin heavy chain gene mapped to chromosomal
region 16q12";
RL Am. J. Med. Genet. 46:61-67(1993).
RN (4)
RP SEQUENCE OF 1093-1972 FROM N.A.
RC TISSUE=Hippocampus;
RA Okajima K.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL
ARTERY, BLADDER, ESOPHAGUS AND TRACHEA.
CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES.
CC -1- PFM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- INVERSION: A CHROMOSOMAL REARRANGEMENT, KNOWN AS PERICENTRIC
INVERSION INV(16)(p13022), PRODUCES A FUSION PROTEIN THAT CONSISTS
OF THE 165 N-TERMINAL RESIDUES OF CBF-BETA (PEPB2) WITH THE TAIL
REGION OF MYH11. THIS REARRANGEMENT IS ASSOCIATED WITH ACUTE
MYELOID LEUKEMIA OF M4EO SUBTYPE.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF001548; AAC31665.1; -

DR EMBL; U91323; AAC35212.1; -

DR EMBL; AB020673; BAA74889.1; -

DR EMBL; D10667; -; NOT ANNOTATED_CDS.

DR EMBL; X69292; CAA49154.1;

DR HSS; P08799; 1MN.

DR MIM; 160745; -

DR InterPro; IPR000048; IQ.

DR InterPro; IPR002928; Myosin_tail.

DR InterPro; IPR002017; Spectrin.

DR InterPro; IPR001609; myosin_head.

DR Pfam; PF00612; IQ; 1.

DR Pfam; PF00063; myosin_head; 1.

DR Pfam; PF01576; myosin_tail; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR ProDom; PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 2.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS00996; IQ; 1.

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;

KW Multigene family; Proto-oncogene; Chromosomal translocation.

FT DOMAIN 1 785

FT MYOSIN HEAD-LIKE.

FT DOMAIN 786 815

FT COILED COIL (POTENTIAL).

FT DOMAIN 844 1934

FT CARBOXYL-TERMINAL.

FT NP_BIND 178 185

FT ATP (POTENTIAL).

FT DOMAIN 661 683

FT ACTIN-BINDING (BY SIMILARITY).

FT DOMAIN 762 776

FT METHYLATION (BY SIMILARITY).

FT MOD_RES 129 129

FT ALKYLATION (SH-1) (POTENTIAL).

FT MOD_RES 701 701

FT ALKYLATION (SH-2) (POTENTIAL).

FT MOD_RES 711 711

FT ALKYLATION (SH-3) (POTENTIAL).

FT CONFLICT 887 889

FT CONFLICT 1263 1266

FT CONFLICT 1558 1558

FT CONFLICT 1610 1611

FT CONFLICT 1786 1786

FT CONFLICT 1938 1938

FT CONFLICT 1972 1972

FT SEQUENCE 1972 AA; 227338 MW; 67665BB2AECE1277 CRC64;

Query Match 3.6%; Score 217; DB 1; Length 1972;

Best Local Similarity 19.8%; Pred No. 0.001;

Matches 243; Conservative 169; Mismatches 441; Indels 374; Gaps 52;

QY 119 KEDAV-----HROSICVYFRADQTKVPEITISIRQAGKIARQELHCPSEFDDTSKKPE 174

DB 857 KEDELQKTRERQ-----KAENLEKEQKHSQLTTEKNLLQEQQAETEL---YAEAE 908

QY 175 VLFPCGRTVAHKKAPALLIDCEIFENHVSGRGSESPRNPHPAAPTQSEPVPRPMK 234

DB 909 M-----RVRLAAKK-----QLEELITHHEWAEARLEEE-----EDRQOQ---LQAEK 946

QY 235 SFSQPL-----RSLAFKELQDGLRSSGFFSFEES-----DIENHLISGH----- 277

DB 947 KAAQMLDLEEQLEEEAARQKLEKVTAEAKIKKLEDEILVMDQNNKLSKERKLL 1006

QY 278 -----NIVQPTDIENTML-----FTIGSEVYLISPDTKKIALEKFNKFCSCSG 325

DB 1007 RISDLTTNLAEEBEAKNLTCLKNKHESMISELEVLKKEKRSQLEKLRKL----- 1060

QY 326 IRHVDHFGICRESSGGGFFHFCYVFOCTNEALVDEIMWTLKQFTVAAVQOTAKAP 385

Db 1061 -----EGDASDFH-----EQIAD-----LQAIQIAELKMLAKKEE 1091

QY 386 LCEGCPLOSLHLKCRIBEGMNSSTKL-BLQKHLTTL-----TQEQATIEE 431

Db 1092 L-----QAALARDDEIAQKNNAKRIEHLGHSIDLOQDLDSERARNKAQKQKDLGE 1146

QY 432 EYQKLRPRNE-----ORENELIISFRLCYEERKQKHEHIGEKQTSQMAA 477

Db 1147 ELEALKTELEDLDTATQOELRAKREQEVTV-LKKALDEETRSHEAQVOEMRQHAQAV 1205

QY 478 ENIGSELPSSATFRFLDMLKNK-----AKRSL-----TESL 508

Db 1206 EELTEQL-EQFKRAKANLDKKNQTLKENADLAGELRVLGQAQOEVEHKKKLEAQVQL 1264

QY 509 ESILSRGNKARGL---OEHSISVDLDS-----SLSSLTNTSKPE 545

Db 1265 QSKSDGERARAEALNDKLVHKLQNEVESVTGMNEAEGKAIKLAKDVASLSSLODQOE-- 1322

QY 546 SVCEKALPISSESPKLLGSSDLSHPEPPAPLSPQOAFRRRANTISHFPIEQ 605

Db 1323 -LQEBETROKLVSTKLQLEBEENS-LQDQLEDE---MEAKONLERHISTLN---IQLS 1374

QY 606 EPPQARGSPGVQSQRKLMRYHSYSTETPHERKDFESKANHLGDSGGTPV-----KTRR 658

Db 1375 D-----SKKKLQDFASTVEALEBGRKRFQKEIENLTQOYEEKAAAYDKLEKTKN 1423

QY 659 HSWRO-----QIFLRVATPQKACDS-----SSRYEDYSELGELPPRSPLE 698

Db 1424 RIQELDLVDLDNORQLVSNLEKKQKQKFDLQALAEKNISSYAD----- 1469

QY 699 PVCDEPGFPPEPKKRTSRELRLWQKAI-----LQQLLLR--MEKNQKLAASENDL 751

Db 1470 -----ERDRAEAREKETKALSARALEEALEAKEELEERTNKLKAEEMDL 1516

QY 752 LNKR--LKLDYEEITPCLKEVTVMKMLSTGSRKIKFDMKMHSAVQGVPRHHRGEI 809

Db 1517 VSKDDVGNVHELEKSKRALETQMEEM-----KTQLELEDELQATEDAKLURLEV 1567

QY 810 WKFLAEQHLKHQFPSKQPKDVPYKELKQTSQOHAIDLGRTPFTPHYFSAQLGAG 869

Db 1568 -----NMQALKQGFERDLQARDEQNEEKRRQLQRLQHEVETEL-----EDERQALAAAA 1618

QY 870 QLSLYNLKAYSLLOEYGVCGGLSFVAGILLHSEEAFAKMLAFMLDFMDGLRKQYRPD 929

Db 1619 KKKLEGLDKLEL-----QADSAIKG-----REDAIKG-----LRK----- 1649

QY 930 MIILQIOMYQLSRLLDHYHR---DLVNHLEHE-IGPSLYAAPWFLTMFASQPLGFVAR 985

Db 1650 ----LQAKMDFORELADARSDEIFATAKENEKAKSLEA-----DLMLQLEDLAAER 1701

QY 986 VEDMIFLQGTVEIVFKVALSLGSHKPLILOHE--NLETIVDFIKSLPLNGLVQMEKTN 1043

Db 1702 ARKQADLEKEELAEELASSLSGRN---ALQDEKRRLEA-----RIAQLBEELE 1746

QY 1044 QVFEMDIAQLQAYEYHVLEELIDSSPLSDNQMDKLEKTNSSLRKNLDLLEQLQV 1103

Db 1747 ---EEQGNMEASDRVRKATQQAQELSNELATERSTAQKNESARQOOLEKRELKSLHE 1803

QY 1104 ANGRQS-----LEATI-----EKLSSSESKLKQAMLTLELERS 1137

Db 1804 MEGAVKSKFKSTIALEAKIAQLEQVEQEAQKQATKSLQKOKKLEILLQVEDERK 1863

QY 1138 ALLQVTEELRRRSAPFSDEPECTOPE 1164

Db 1864 MAEQYKEQAQKGNARVKQLKQLEAE 1890

RESULT 9

MYHB_CHICK

ID MYHB_CHICK STANDARD; PRT; 1978 AA.

AC P10587;

DT 01-JUL-1989 (Rel. 11, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)
 Myosin heavy chain, gizzard smooth muscle.
 Gallus gallus (Chicken).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 Gallus.
 NCBI_TaxID=9031;
 [1]
 SEQUENCE FROM N.A. PubMed=3892941;
 MEDLINE=88118916; Hamada Y., Katsuragawa Y., Imamura M., Mikawa T.,
 Masaki T.;
 Yanagisawa M., Hamada Y., Katsuragawa Y., Imamura M., Mikawa T.,
 Masaki T.;
 "Complete primary structure of vertebrate smooth muscle myosin heavy
 chain deduced from its complementary DNA sequence. Implications on
 topography and function of myosin."
 J. Mol. Biol. 198;143-157(1987).
 [2]
 REVISIONS.
 Masaki T.;
 Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE OF 1-203.
 MEDLINE=88032919; PubMed=3312184;
 Maita T., Onishi H., Yajima E., Matsuda G.;
 "Amino acid sequence of the amino-terminal 24 kDa fragment of the
 heavy chain of chicken gizzard myosin."
 J. Biochem. 102;133-145(1987).
 [4]
 X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-818, AND REVISIONS 204-215.
 MEDLINE=98412652; PubMed=9741621;
 Dominguez R., Freydon Y., Trybus K.M., Cohen C.;
 "Crystal structure of a vertebrate smooth muscle myosin motor domain
 and its complex with the essential light chain: visualization of the
 pre-power stroke state."
 Cell 94;559-571(1998).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.

 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; X06546; CAA29793.1;
 PIR; S03166; S03166.
 PDB; 1BR1; 09-SEP-98.
 PDB; 1BR2; 09-SEP-98.
 PDB; 1BR4; 09-SEP-98.
 InterPro; IPR000048; IQ.
 InterPro; IPR004009; Myosin_N.
 InterPro; IPR002928; Myosin_tail.
 InterPro; IPR002017; Spectrin.
 InterPro; IPR001609; myosin_head.
 Pfam; PF00612; IQ; 1.
 Pfam; PF00063; myosin_head; 1.
 Pfam; PF02736; Myosin_N; 1.

DR Pfam: PF01576; Myosin_tail; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRODOM: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS00096; IQ; 1.
 KW Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Calmodulin-binding;
 KW Multigene family; 3D-structure.
 FT INIT_MET 0
 FT DOMAIN 1 790 MYOSIN HEAD-LIKE.
 FT DOMAIN 791 820 IQ.
 FT DOMAIN 849 1978 RODLIKE TAIL (S2 AND LHM DOMAINS).
 FT DOMAIN 849 1978 COILED COIL (POTENTIAL).
 FT NP_BIND 176 183 ATP.
 FT DOMAIN 666 688 ACTIN-BINDING.
 FT DOMAIN 767 781 ACTIN-BINDING.
 FT MOD_RES 1 1 BLOCKED.
 FT MOD_RES 127 127 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 706 706 ALKYLATION (SH-1).
 FT MOD_RES 716 716 ALKYLATION (SH-2).
 FT CONFLICT 127 127 MISSING (IN REF. 3).
 FT CONFLICT 204 215 KDTISITQPSFS -> RTPSLKVLFP (IN REF. 1).
 SQ SEQUENCE 1978 AA; 228663 MW; B7B6C923E5273D93 CRC64;
 Query Match 3.5%; Score 212.5; DB 1; Length 1978;
 Best Local Similarity 19.2%; Pred. No. 0.0017;
 Matches 191; Conservative 147; Mismatches 310; Indels 345; Gaps 43;
 QY 356 NEALVDIEMTLKQA-----FTVAAGV-OTAKAPAOQACGCP 392
 DB 1037 HESMISELEVLKKEEKSQLEKIKKLEGESDLHQIAELQIAELAKLAQ 1094
 QY 393 QSLHKLCEIRIEGMSKTKL-----ELQKHLTTL-----TNQQAIFEEVQ 434
 DB 1095 EELQAALARLEDEFSQNNALKKIRELESISDLQEDLESEKAARNKAQKRDLESE 1154
 QY 435 KLRPNE-----ORENELISFLRCYBEKQKEHIGEMKQTSMAANI 480
 DB 1155 ALKTELEDLTDTTQOELRAKREQVTV-LKRALEEETRTHEAQVQEMRKHTQAVEL 1213
 QY 481 GSELP-SATFRFLMLKNAK-----RSITES-----LESIL 512
 DB 1214 TEQLEQFKRAKANDKTKQTLKONADIANEIRLSQAKQDVHEKKKLEVLQDLQSKY 1273
 QY 513 SRGNKAR-GLQF--HSISVDLSSLSSTLSNTSKEPSYCEKEALPI-----SE 557
 DB 1274 SDGERVTELNEKVHKLQIEVE-NVTSLLNEAESKNIKLTKDVATLGSQLODTQELLQEE 1332
 QY 558 SSFKL-----LGSSSEDLSSSESHLPPEPAPLSPQOQAFRRRANTLSHPICQPPQPAR 612
 DB 1333 TROKLVNTTKLRQLEDKNSLQEQLEDE--VEAKONLERHISTLT---IQLSD----- 1380
 QY 613 GSPGVSRKLMRYHSVSTETPERK-----DFESKANHLGDSGGTTPVKTRRH- 659
 DB 1381 -----SKKKQEF-TAVVETMEEGKKLQRIEISLTQOFEEKAASYDKLETKRKLQEL 1434
 QY 660 -----SWRQIFLRVATPQKADS-----SSRYEDYSELGELPPRSPLEPVCBDG 704
 DB 1435 DDLVVDLDNQRLVSNLEKKQKFDQMLAEKNISSYAD----- 1474
 QY 705 PFGPPPEKKRTSRELRLWQKAI-----LQOILLR--MEKENQKLOASENDLLNKR-- 755
 DB 1475 -----ERDRAEAAREKETKALSARLEALEAKEELERTNMKKAEMEDLVSSKDD 1527
 QY 756 -----LKLDEYEITPCLEKVTTWKMLSTPGRSKIKFKDMEKHSVAGGV 801
 DB 1528 VGKNVHELEKSKRTLEQQVEEMKTOLEED--ELQAAEDAKLRLVYNQAM----- 1577
 QY 802 PRHRGEIWKFLAEQFHLKHQFPSPKQPKDVYKELKQLTSQOQHALLIDLGRTFPHY 861
 DB 1578 -----KSQFERDLQARDQNEKRRQLKQLHEHETEL-----EDERK 1615

QY 862 FSAQLGAGQLSLYNIKAYSLDQEVGYCOGLSFVAGIILLHMSBEEAFKMLKFLMFDWG 921
 Db 1616 QRALAAAKKLEVDYKD--LESQVDSA--NKAREBAIKQLRKLAQAM- 1659
 QY 922 LRKOYFDM--IILQIQMYQLSRLLDYHRLDLYN-HLEEH 959
 Db 1660 --KDYQDLDLDAAREEIEFATARENEKKAKNLEALQLOEDLAARARKQADLEKEE 1717
 QY 960 IGPLSYAAPWFLTMFASQPLGVARFDMIFLQGTVEVIFKVALSLGSHKPLIIOHNL 1019
 Db 1718 MABELASANSGRTS--LQDEKRRLREARIAQL--EEELDEHSNI 1757
 QY 1020 ETIVDFIKSLPNLGLVQMEKTNQVFMENDIAKQLQAYEYHVLOEELIDSSPLSDNQR 1079
 Db 1758 ETMSD--RMRKAVQQA--EOLNNELATERAT 1784
 QY 1080 MDKLEKNTSLRKQNLDDLEQLQVANGRIOS-----LE---ATIEKLLSSSEKKAQ 1128
 Db 1785 AOKNENARQOLERQNKRLSKLQEMEGAVKSKFKSTIAALEAKIASLEQLQEAQAKQA 1844
 QY 1129 ML-TLELE-----RSALLQTVVEELRRRSAPSDR 1156
 Db 1845 AAKTLRQKDKKLKDALLO-VEDERKQAEQYKDQ 1876

RESULT 10

GOG4_HUMAN STANDARD; PRT; 2230 AA.
 AC Q13439; Q14436; Q13270; Q13654;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa
 DE golgin) (Golgin-245) (72.1 protein).
 GN GOLGA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96215236; PubMed-8626529;
 RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
 RT "Molecular characterization of trans-Golgi p230: a human peripheral
 RT membrane protein encoded by a gene on chromosome 6p12-22 contains
 RT extensive coiled-coil alpha-helical domains and a granin motif.";
 RL J. Biol. Chem. 271:8328-8337(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Seelig H.P.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 131-2230 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE-96125112; PubMed-8537393;
 RA Fritzel M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
 RT "Molecular characterization of golgin-245, a novel Golgi complex
 RT protein containing a granin signature.";
 RL J. Biol. Chem. 270:31262-31268(1995).
 RN [4]
 RP SEQUENCE OF 524-672 FROM N.A.
 RC TISSUE-Gastric fundus;
 RA Balague C.;
 RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
 CC -1- FUNCTION: MAY PLAY A ROLE IN VESICULAR TRANSPORT FROM THE TRANS-
 CC GOLGI.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
 CC GOLGI MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS MAY ARISE BY ALTERNATIVE
 CC SPLICING.
 CC -1- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND
 CC IN HEPATITIS B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U41740; AAC50434.1; -
 DR EMBL; X82834; CAA58041.1; -
 DR EMBL; U31906; AAC51791.1; -
 DR EMBL; X76942; CAA54261.1; -
 DR MIM; 602509; -
 DR InterPro; IPR000237; GRIP.
 DR Pfam; PF01465; GRIP; 1.
 KW Golgi stack; Antigen; Coiled coll; Alternative splicing.
 FT DOMAIN 133 237
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 276 1011
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1033 1214
 FT COILED COIL (POTENTIAL).
 FT VARSPLIC 2154 2185
 FT TPYKGNLYHTVDSLEGTEFEYLRKLVFEY -> HLTKV
 FT AICTRMSHSLNPLNSICERKCLSI (IN ISOFORM
 FT 2).
 FT MISSING (IN ISOFORM 2).
 FT MISSING (IN ISOFORM 3).
 FT FTSPRSGLF -> SWLRSS (IN ISOFORM 4).
 FT R -> K (IN REF. 3).
 FT Y -> H (IN REF. 3).
 FT T -> A (IN REF. 3).
 FT K -> E (IN REF. 3).
 FT T -> A (IN REF. 3).
 FT K -> E (IN REF. 3).
 FT K -> N (IN REF. 3).
 FT CONFLICT 682 682
 SQ SEQUENCE 2230 AA; 261139 MW; 3BB733DBEA86134 CRC64;
 Query Match 3.4%; Score 208; DB 1; Length 2230;
 Best Local Similarity 19.9%; Pred. No. 0.0035;
 Matches 178; Conservative 150; Mismatches 309; Indels 256; Gaps 41;
 QY 372 TVAAVOOTAKAPALCEGC--PLQSLHKICERI-----EGMNSKYKLEQLKHLTTL 421
 Db 280 TLETLQQRVKRQBNLKRCKETIQSHKEQCTLLTSEKALQEOQLDERLOLEKIKDLHMA 339
 QY 422 TNOEQATIEFEVKLRPRNEORENELIISFLRCLYEKKEHIGEMK-OTSMAAENI 480
 Db 340 ETKLITQIRDAKNLIEQLEODKGWIAETKQRMHETLEMKEEIAQLSRKQMTQ-- 397
 QY 481 GSELPPSATRFRDLMLNKAKRSLTESLSILSRGNKARGLQEHSHISVDLSSLSLTLN 540
 Db 398 GEELREQ-----KEKSERAAPELEKALSTAQKTEAR-RKLKAEMDEQI-KTIEK 446
 QY 541 TSKEPSVCEKEALPISESSFLLGSSSEDLSSSESHLPEEPAPLSPQQAFFRRANTLSHF 600
 Db 447 TSEERISLQDEL-----SRVKQEVVDVMKKSSEQIA----- 479
 QY 601 PIECQPPPARCGSPGVSRQKLMRYHVSVEPHERKDFESKANHLGDSGTPVKTRRHS 660
 Db 480 -----KLQKLH-----EKELARKEQLTKK-----LQTRERE 506
 QY 661 WRQOIFLRVATPOKACDSSSRVEDYSELGELPPRSLEPVCEDGPFQPPPEKKR----- 715
 Db 507 FQEQ--MKVALEK-----SQSEYLIKISQKEQESLAEEL-----ELQKKAILTE 550
 QY 716 TSRELRELQKAILQOILLRMEKENQK-LQASENDLLNKRKLKDYEEETPCPKVTTVM 774
 Db 551 SENKRLDQEAETRYTRILELESLSLEKSLQENKQSKDLAVHLEAK-NKNKEITVMV 609
 QY 775 EKMSTPGRSKIKFDEKMKHSAVGQGVPRHHRGEIKWFLAEQPH-LKHQPPSK----- 826
 Db 610 EXH-----KTELES-----KHQDALM---TEKLOVLQOQYQTEMKLEKREK 648

QY 827 -OOPKDPYKELLKOLTISOHAILDLGRFTFTHPVFSAQAGQLSLNLIKAYSLLDQ 885
 Db 649 CQEKETLLKD--KELIQAHEEWN-EKLEKLDVKQTELESLSSEVLKARHKLEE 705
 QY 886 FVGYCOGLSFVAGILLHMSSEAFKMLKFLMDMGLRK---QYRPMII----- 932
 Db 706 E-----LSVLKQDTDKMKQLEAKMDEQKNHHQOQQVDSIIKEHVSIOQT 750
 QY 933 ---LOIQMYQLSRLLDHYRDLYNHLBEHEI-----GFSLYAAPFWLTFWAS- 976
 Db 751 EKALDKQINOELLKKE--RD--KRLKEQAHVENLEADIKRSEGELOQASAKLDVFOY 806
 QY 977 QPPLGFVARVFMIFLQGTVEVLFKVALSLGSHKPLIL-----OHENLETIVDFI 1026
 Db 807 QSATHEQKAYE-----EQAQLOQLDLLETERILLTKOVAEVAQKDVCTELDAH 859
 QY 1027 KSTLNLGLVQMEKTIQVNFENDIAQLOQAYEYHVILQEEELIDSPISLDNQMDKLEKT 1086
 Db 860 KIOVQDL-MQOLEKQ-NSEMEQKVKSLTQVYE-----SKLEDGNKEQEQTKQILVEKE 910
 QY 1087 NSSL-----RKQNLDLLEQ-----LOVANGRIQISLEATI 1115
 Db 911 NMILQMRGQKKEIILTKQSAKEDSIHLNEEYETKFKNOEKKMEYKQAKEMQETL 970
 QY 1116 -EKLSSSESKLQOAMLTLELERSALLQTVEEELRRSABPSDREPECTOPEPTG 1167
 Db 971 KKLLDQEAKLK-----ELENTAL-----ELSQREKQFNKMLEMAQANSAG 1013

RESULT 11
 CENE_HUMAN
 ID CENE_HUMAN STANDARD; PRT; 2663 AA.
 AC 002224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENP-E.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.-J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis.";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.-J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor.";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.-J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC -|- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -|- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
 CC -|- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -|- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL; Z15005; CAA78727.1; .
 DR PIR; S28261; S28261.
 DR HSSP; P17119; 3KAR.
 DR MIM; I17143; .
 DR InterPro: IPR001752; kinesin.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335 KINESIN-MOTOR.
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP_BIND 86 93 ATP (BY SIMILARITY).
 SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8C8 CRC64;
 Query Match 3.4%; Score 207; DB 1; Length 2663;
 Best Local Similarity 17.8%; Pred. No. 0.005;
 Matches 160; Conservative 165; Mismatches 290; Indels 284; Gaps 34;
 QY 359 LVDEIMTLKQAFVAAVQQTAKAPALCEGCPLQ-----S 394
 Db 913 ITEKLOQTLEEVKTLTQEKDLDL---QLOESLQIERDQLKSDIHDVTNNIDTQSLRNA 969
 QY 395 LHKLCERTEGMNSKTL--ELQKHLTLTN-----QEQAIFEEYVKLRPRNEQ-- 442
 Db 970 LESLKHQOETINTLASKISEEVSRNLHMEENTGETKDFQOKMVGIDKKDLEAKNTQL 1029
 QY 443 ---RENELITSFLRCLYEKQKEHIGEMKQTSQMAAENSGELSPSATFRDLMLKN 498
 Db 1030 TADVKNDEII-----EQQRKIFSLQEKNEQLQOMLESVIAE-----KE 1067
 QY 499 KAKRSLTESLSILSRGNKARGL-----QHSISVDLDSLSSTLSNTSKEPS 546
 Db 1068 QKTDLKENIEMTNIENQELRLGLDELAKQOQIVAQERHAIKKEGELSRDCDLAEVEE 1127
 QY 547 VCEKALPISSESKLLGSSDLS-----DSHSLPEEPAPLSPQQAFFRRANTLSHP 601
 Db 1128 KLKESQQLQEQKQOLLNVQEMSEMOKKINEIENKNE-----LKNKELTLEHME 1178
 QY 602 IECQEPPOPARGSPCVSORKMLRYHSVSTPHER-----KDFESKANHL-----GDSGG 651
 Db 1179 TERLE-----LAQKNENYEEVKSITKERVKLQKSFETERDHLRGYIREIA 1228
 QY 652 TPVKTRRHSWRQQIFLRVATPQKACDSSSRVEDYSELGELPPSPLEPCVCEGPGPPPE 711
 Db 1229 TGLQTKELKTAHILK----- 1245
 QY 712 EKKRTSRELRLWQKAILQIQLLLMRKENOKLOASENDLLNKRKLQDYEEITPCLKVEY 771
 Db 1246 EHQETIDELRRSVSEKTAQIINTQDLKSHTKLQ-EIIPVLHEE-----QELLPNVKVS 1299
 QY 772 TWKEM-----LSTPGRSKIKFDMKMSHSAVQGVPRHHRGEIWKFLAOFHLKHQFSPK 826
 Db 1300 ETQETMNELELLTEQSTTKDSTTLARIE-----MERLNLKFFQES 1340
 QY 827 QQPKDVPYKELLKOLTISOHAILDLGRFTFTHPVFSAQAGQLSLNLIKAYSLLDQ 886
 Db 1341 Q-----EIKSUTKERNLKTKEALEVHKDOLKEHI---RETLAKIQESQSQQS 1389
 QY 887 VGYCOGLSFVAGILLHMSSE--EAFKMLKFLMDFMDGLRKQYRPD---MILQIQMOLS 941

Db 1390 -----LNMEKDNETHIVSEM-----EQFKPKDSALLRIEIMLGSL 1427

QY 942 RLLHDYRDLYNHLSEHEIGPSLYAAPWFLTFASFQPLGFVARVDFMIFLQGTVEIFKV 1001

Db 1428 KRLQESHDEMKVAKED-----DLQRLQ--EVLQSE 1457

QY 1002 ALSGLSHKPLILQHNLETFIVDFIKSTLPNLGLVQMEKTIQOV-----FEMDIA---K 1052

Db 1458 SDQLKENIKRIVAKH--LETEEL--KVAHCCKLKEQETINELRVNLSEKETEISTIQ 1512

QY 1053 OLQA-----YEYEHVQLBELDSSPLSDNQMDKLEKTNSSLRKQNLDLLE- 1099

Db 1513 QLEAINDKLNKQTOEYIEKEQELNIKOI---SEVOENVELKQFKHRKAKDSALQSIES 1569

QY 1100 -QLQVANGRIQSEATIEKLLSSKSLKQAMLTLELERSALLQTVLELRRRSAPSDRE 1157

Db 1570 KMLELN-RLQESQEEIQIMKEEMKRVQEAQLQIEROQLKENTKEIVAKMKESQEKE 1627

RESULT 12

MYHB_RABIT

ID MYHB_RABIT STANDARD; PRT; 1972 AA.

AC P35748:

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Myosin heavy chain, smooth muscle isoform (SMHHC).

GN MYH11.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_Taxid=9986;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92073350; PubMed=1961735;

RA Babji P., Kelly C., Periasamy M.;

RT "Characterization of a mammalian smooth muscle myosin heavy-chain

RT gene: complete nucleotide and protein coding sequence and analysis of

RT the 5' end of the gene.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).

CC 1- FUNCTION: MUSCLE CONTRACTION.

CC 1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC 1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC 1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC 1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY

CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC 1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

CC SUBFRAGMENT (S2).

CC 1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC 1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: M77812; AAA31395.1; --

DR FIR: A41604; A41604.

DR HSSP: P08799; LMMD.

DR InterPro: IPR000048; IQ.

DR InterPro: IPR004009; Myosin.N.

DR InterPro: IPR002928; Myosin.Tail.

DR InterPro: IPR002017; Spectrin.

DR InterPro: IPR001609; myosin_head.

DR Pfam: PF00612; IQ; 1.

DR Pfam: PF00063; myosin_head; 1.

DR Pfam: PF02736; Myosin.N; 1.

DR Pfam: PF01576; Myosin.Tail; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR ProDom; PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS00096; IQ; 1.

KW Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;

KW Multigene family.

FT DOMAIN 1 784 MYOSIN HEAD-LIKE.

FT DOMAIN 785 807 IQ.

FT DOMAIN 844 1934 COILED COIL (POTENTIAL).

FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.

FT NP_BIND 178 185 ATP (POTENTIAL).

FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).

FT DOMAIN 763 777 ACTIN-BINDING (BY SIMILARITY).

FT MOD_RES 129 129 METHYLATION (SH-1) (POTENTIAL).

FT MOD_RES 701 701 ALKYLATION (SH-2) (POTENTIAL).

FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).

SQ SEQUENCE 1972 AA; 227318 MW; 2061A224288D6A4C CRC64;

Query Match 3.3%; Score 202; DB 1; Length 1972;

Best Local Similarity 19.8%; Pred. No. 0.0062;

Matches 187; Conservative 127; Mismatches 352; Indels 278; Gaps 37;

QY 377 QQTAKAPALCE-----GCPLASLHLKCRIEGMNSKTKL-----ELQKHLTTL----- 421

Db 1069 EQIADIQAQIAELKMLQAKKEELQAALAELETSQKNALKRIELEGHISDLQEDLD 1128

QY 422 -----TNOQATIFEVVKLRPRNE-----ORENELIISFLRCLYEEK 459

Db 1129 SERAARKAEKQRKLDGELEALKTELEDTLTATQOELRAKREVEVTV-LKKALDEET 1187

QY 460 QREHIHIGEMKQTSMAAENIGSELPP-SATFRFLDMLK-----NKAQ 501

Db 1188 RSHEAQVQEMRKHTQVVEELTEQLQFKRAKANLTKQTLEKENADLAGELRVLGQAK 1247

QY 502 RSL-----TESLSILSRGNKARGL---QEHISIVDLDS----- 532

Db 1248 QEVEHKKKLEVLQQLQSKSDGERARAEALNDKVKHLQNEVESVTGMLSEABGKAUKLA 1307

QY 533 ----SLSSTLSNTSKEPSVCEKEALPISESSFKLGSSDSDSHLPEEPAPLSPOQ 588

Db 1308 KEVASLGSLQDTQELLQETROKLVSTK----LQLEDERNSLQEQLEDEE---MEAKQ 1360

QY 589 AFRRRANTLSHFPIECQEPQPARGSPGVSGRKLMIYHSV--STETPHER--KDFESKAN 644

Db 1361 NLERHISTLN---IQLSD-----SKKKLQDFASTVESLEBGRKRFQKEISLTQ 1406

QY 645 HLGDSGGTPVKVRRHSWR-----QQIFLRVATPQKACDS-----SSR 681

Db 1407 QYEEAAAYDKLEKTKNRLQQLQELDLVDLDNQRLVSNLEKKQKFKDQLLAEKNISSK 1466

QY 682 YEDYSELGELPPRSPLEPCVEDGPFPPPEKKRTSRRELRLWQKAI-----LQOILLR 736

Db 1467 YAD-----ERDRAEAAREKETKALSARALEALEAK 1499

QY 737 --MEKENQKQASENDLANKR--LKLDYBEITPCLKEVTVVWEKMLUSTPGRSRIKFDMEK 792

Db 1500 EELERTNKLKAAEMEDLVSSKDDVGKGVHLEKSKRALETQMEEM-----KQLEEE 1550

QY 793 MHSVAGQGVPRHRRGIWFKFLAEQFHLKHQFSPKQPKDVPVYKELLKQLTSQOHAIDL 852

Db 1551 LEDELQATEDAKRLLEV-----NMQALKVQFERDLQARDEQNEEKKRRQRLQRLHETEL 1605

QY 853 GRTFPTHYFSAQLGAGQLSYLNILKAYSLLDQEVGCGQLSFVAGILLHMHSEEAFAKM 912

Db 1606 ----EDERKQRLAAAKKKLGDLKDEL-----QADSAIKG-----REEAIKQ 1646

QY 913 LKFLMFDMLGRKQYRPMIIIOIOMYQSLRLHHDYHR---DLYNHLBEHE-IGPSLYAAP 968
 Db 1647 L-----LKIQAQMKDFQRELEADARASDRDEFATFATKENERKAKSLEA-- 1687
 QY 969 WFLTMFASQFPLGFVARVEDMIFLQGVTEFKVALSLGSHKPLILOHE--NLETIVDFI 1026
 Db 1698 ---DLMQLQEDLAAEATARKQADLEKELABELASSISGRN---ALQDEKRRLEA----- 1736
 QY 1027 KSTLPNGLVOMEKTIINOVFENDIAKQIQAYEYHYHVLQEEELIDSSPLSDNORMDKLEKT 1086
 Db 1737 -----RIAQLEEELE---EEQCNMEAMSDRVKATQQAQSLNELATERSTAKNESA 1786
 QY 1087 NSSLRKNLDLLEQLQVANGRIQS-----LEATI-----EKLKS 1120
 Db 1787 RQQLERONKELSKLQMEGAVKSKFKSTIAALEAKIAQLEEQVEQAREKQAQAALKQ 1846
 QY 1121 SESKLQAMLTLELERSALQTVBELRRRSAPESDRPECTQPE 1164
 Db 1847 RDKKLKEMLLQVEDERKMAEQYKEQAEKGNKAKVKOLKQLEAE 1890
 RESULT 13
 MYHA_BOVIN STANDARD; PRT; 1976 AA.
 AC Q2791;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
 DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
 GN MYH10.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ohara M., Ishiguro N., Shinagawa M.;
 RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 204-302 FROM N.A.
 RC TISSUE=Brain cortex;
 RX MEDLINE=95301542; PubMed=7782316;
 RA Itoh K., Adelstein R.S.;
 RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle
 RT myosin heavy chain II-B.";
 RL J. Biol. Chem. 270:14533-14540(1995).
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CC CAPPING (BY SIMILARITY).
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB022023; BAA36494.1;
 CC EMBL; U15716; AAA87715.1;
 CC HSSP; P08799; 11VK
 DR InterPro; IPR000048; IQ.

DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 KW Myosin: ATP-binding; Calmodulin-binding; Actin-binding;
 KW Coiled coil; Alkylation; Multigene family.
 FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
 FT DOMAIN 786 815 IQ.
 FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
 FT NP_BIND 178 185 ATP (POTENTIAL).
 FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1976 AA; 229097 MW; 6144354451C0F790 CRC64;
 Query Match 3.38; Score 197; DB 1; Length 1976;
 Best local similarity 19.58; Pred. No. 0.011;
 Matches 193; Conservative 127; Mismatches 318; Indels 354; Gaps 39;
 QY 357 EALVDEIMTMLKQAFVAAVQOTAKAPALCEGCPL-----OSLHK----- 397
 Db 1075 QAQIDELKI-----QVAKKEEL-QGALARGDDETLHNNALKVVRLOAQIA 1121
 QY 398 -LCBRIEGMNSKTKLELQK-----HTLTNQEQATIFEV--OKLPRNQRENELII 449
 Db 1122 ELQEDFSEKASRNKAEKQKRDLSLEALKTELEDLTDTAAQQLTKRKEQVAEL-- 1179
 QY 450 SFLCLYEEKQKEHIHIGEMKQTSQMAAENIGSELPSATFRDLMLKKA----- 500
 Db 1180 --KKALEETKSHBAQIQDMQRHATALEELSEOL-EQAKRFKANLEKNKOGLETDNKL 1236
 QY 501 -----KR-----SLTESLESILSRGNKARGLOEHSISVDL---DSSLST 537
 Db 1237 ACEVKVLQVKAESEHKRKLDAQVQELHAKVSEGDRLR-----VELAEKANKLQNE 1288
 QY 538 LSNWTSKPSVCEKEALPISSEPKLLGSSDLSDESHLPPEEPAPLSPQQAFFRRNTL 597
 Db 1289 LDNVSTLLEAEKKGIRKFAKDAAGL---ESQLQDTQELQEEET-----ROKLNLS 1335
 QY 598 SHRP-----TECOEPPOPARGSPGVORSOKLMRYHVSSTETPHE----- 635
 Db 1336 SRIRQLSEERSLSQEQEEEEEARRS---LEKQLAQALQALDTYKKVDDDLGTIENLEE 1392
 QY 636 -----RKDFESKANHLGDSGGT-----PVKTRRH-----SWRQOIFLRVATPQKA 675
 Db 1393 AKKLLKDVLSVLSQLEEKALAYDKLEKTIRLQQLQELDDLVDLDHQRIVSNLEKKQK 1452
 QY 676 CDS-----SSRYEYSELGELPPRSPLEPCVDCGDFGPPPEKKRTSELRELWQKA 727
 Db 1453 FDQLLAEBEKNISARY-----AEERDRAEAAREKETKA 1485
 QY 728 I-----LQIILLR--MEKENOKLQASENDLLNKR-----LKLDYEET 764
 Db 1486 LSLARALEALEAREAEARNQKQURADMEDLMSSKDDVGKKNVHELEKSKRALEQVEEMR 1545
 QY 765 PCLKEVTVMEKMLSTPGRSKIKFDMKMHSAVGQGVPRHHRGEITWKFLABQFHLKHQFP 824
 Db 1546 TQLEED--ELQATEDAKLRLEVNMQAM-----KAQFE 1577
 QY 825 SKQPKDVPYKELKQITSQOHAILEDGRFTPTHYPSAQLGAGQSLYLKAYSLLD 884
 Db 1578 RDLQTRDEQNEEKRLIKQVRELEAELEDERKORALAVASKKMEIDKDL----- 1629
 QY 885 QEVGYCQGLSFVAGILLHMHSEEAFFKMLFLMDFMDGLRKQYRPMIIOIOMYQSLRL 944

MYH9_HUMAN
 ID MYH9_HUMAN STANDARD; PRT; 1960 AA.
 AC P35579; G08085;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE MYOSIN heavy chain, nonmuscle type A (Cellular myosin heavy chain,
 DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
 GN MYH9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
 RA Bagguely C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
 RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
 RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
 RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
 RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
 RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
 RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
 RA Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
 RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kersey J.,
 RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
 RA Lloyd C., Lloyd D.M., Marlyn I.D., Mashreghi-Mohammadi M.,
 RA Matthews L., McLean O.T., McLeay J., McLaren S., McMurray A.,
 RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
 RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
 RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
 RA Snalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
 RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
 RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
 RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
 RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
 RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
 RA Mitsuyma S., Roe B.A., Chang F., Chu L., Crabtree J., Deschamps S.,
 RA Do A., Dorman A., Fan F., Fu Y., Hu P., Hua A., Kenton S.,
 RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.P., Loh P., Mala E.,
 RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
 RA Sloan D., Song L., Wang Q., Wang Y., White J., Willingham D.,
 RA Wu H., Yao Z., Zhan M., Zhang G., Chisoe S., Murray J., Miller N.,
 RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
 RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
 RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
 RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K.,
 RA Nelson J., Korf I., Bedell U.A., Hillier L., Mardis E., Waterston R.,
 RA Willson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tiliakou Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [2]
 RP SEQUENCE OF 1-1337 FROM N.A.
 RX MEDLINE=92003925; PubMed=1912569;
 RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
 RA Arnaout M.A., Clayton L.K., Tenen D.G.;
 RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
 RT cDNA clones, characterization of the protein, chromosomal
 RT localization, and upregulation during myeloid differentiation.";
 RL Blood 78:1826-1833(1991).
 RN [3]
 RP SEQUENCE OF 1-715 FROM N.A.
 RX MEDLINE=91316803; PubMed=18660190;
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
 RA Guila D., Adelstein R.S., Weir L.;
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located
 on different chromosomes";
 RL Circ. Res. 69:530-539(1991).
 RN [4]
 RP SEQUENCE OF 714-1960 FROM N.A.
 RX MEDLINE=90138958; PubMed=1967836;
 RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
 RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
 RT through alternative polyadenylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
 RN [5]
 RP VARIANT DFNA17 HIS-705.
 RX MEDLINE=20489856; PubMed=11023810;
 RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Miatre A.N.;
 RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
 RT nonmuscle myosin MYH9.";
 RL Am. J. Hum. Genet. 67:1121-1128(2000).
 RN [6]
 RP VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.
 RX MEDLINE=20428192; PubMed=10973259;
 RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
 RA Grigori G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
 RA Iolascon A., Zelante L.L., Savoia A., Balduini C.L., Noris P.,
 RA Magrini U., Belletti S., Heath K.E., Babcock M., Gluckman M.J.,
 RA Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:103-105(2000).
 RN [7]
 RP VARIANTS MHA ILE-1155 AND LYS-1841.
 RX MEDLINE=20428193; PubMed=10973260;
 RA Kelley M.J., Javien W., Ortel T.L., Korczak J.F.;
 RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
 RT May-Hegglin anomaly.";
 RL Nat. Genet. 26:106-108(2000).
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CC CAPPING.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
 CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME
 CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC -1- WITH ADDITIONAL ALPOPT-LIKE CLINICAL FEATURES OF SENSORINEURAL
 CC DEAFNESS, CATARACTS AND NEPHRITIS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
 CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS
 CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
 CC COCHLEOSACCULAR DEGENERATION.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 282215; CAB05105.1; -
 CC EMBL; M81105; AAA59888.1; -
 CC EMBL; M69180; AAA61765.1; -
 DR
 DR

[illegible]

Db	1234	KVLLQKGDSEHKKR	KVEAQLOEQVKTEG	ERVNTEADKVTKLOVELDNVTGL	SSQSD	1293								
Qy	532	-----SSLST	SLNTSKEPSC	KEALPIS-----ESSFK	ILGSGE	572								
Db	1294	SKSSKITD	FOFSALE	SOLODTQELLQ	ENRQKLSLTKLQ	VEDEKNSF-----EQLEEE	1348							
Qy	573	SES--HLPE	PAPLSPQAFRR	RANTLGHF	IEOQPPQ	PARGS-----PGVSQ	---KL	632						
Db	1349	EEAKHNLE	QKQIATLHAQ	VADMKK	MMEDS---VGC	LETAEEVKR	KLQDLQGLSG	SHOEKV	1405					
Qy	623	MRY--HSV	STPHERKDFES	KANHLG	SGGTPVKTKRRH	SRQOIFLR	VATPOK	ACDSS	679					
Db	1406	AAYDKLEK	TKRLQEL	DDLVL	VDLQH-----Q	RGSAC	NLE	1440						
Qy	680	SRYEDY	SELGELPP	RSPLEPC	VEDGPF	GGP-PEEK	KRTSREL	RELWQAILQ	ILLRME	738				
Db	1441	KKQKFDQ	-----LAE	KTISAKYA	EBDR	AEAEARE	KETKA-----LS	LARA	1484					
Qy	739	KENQK	LOASEN	LLNNKRL	KLYEETPC	LKVTTW	EKMLSTP	GRS	KIKFDM	EKMHSAVG	798			
Db	1485	LEBAMQ	KAELE	RNLNKF	QFTMEDL-----M	SKDDG	VKS	VELEK	SRKRALE	1531				
Qy	799	QGVPRH	RHGEIN	KFLAE	QPH-----L	KHPFS	QKQPKD	VPYK	ELLK	839				
Db	1532	QQV-----E	EMK	TQLEEL	DELQATEDAK	LRLEVN	LQAMKA	QFERD	LQGRD	EQSEKKK	1585			
Qy	840	QVTSQHA	ILIDL	GRTPPH	YPFSQA	LQAGOL	SLYNIL	KAYSL	LDQEV	GYCQGLS	FVAGI	899		
Db	1586	QLVRQV	REM-----E	AELDER	KQRSM	VAAR	KKLEMDL-----K	DLEAHI	1626					
Qy	900	LLAHMSE	EA	KMLKFL	MFDMGL	RQYR	PDMLII	QIOM	YOLS	RLHLD-----Y	HRDLYNHLE	956		
Db	1627	DSANKR	DEA	IKQ-----L	RK-----L	QAKM	KDCM	RELD	DT	RASREI	LAQAK	1669		
Qy	957	EHEIG	PSLYA	APWFL	TMFASQ	PLGF	VARV	DMIF	LOQTE	VI	FKVAL	SLGLSHK	PLILOH	1016
Db	1670	ENEK	LKLSMEA-----E	MIQ	LQ-----E	LA	AAERAK	RQAK	1700					
Qy	1017	ENETIV	DIKSTL	PNLNL	VQMEK	TINQ	VFEMD	IAKOL	QAEV	EYHV	LQEL	DDSS	PLSD	1076
Db	1701	QERDE	LAD	EIANSS	GKGAL	ALEEK---R	RLS	ARIA-Q	LEF-EL	SEEG	ONT	ELIN	DRUKKA	1755
Qy	1077	NORM	KDL-----E	KTNSS	RKRN	LDLLE	QLO	VANG	RIGOS-----L	EA	T	1114		
Db	1756	NLQIDQ	INTDLN	LSHQA	KNENAR	QQLER	QNKEL	KVYK	LOEM	E	GT	VKS	KASITALEAK	1815
Qy	1115	I---EKL	SESEK-----L	QAMU	TLE	RSAL	QTV	VEEL	RRRS	AE	1152			
Db	1816	IAQLE	QOLNET	K	QAAKQ	QVRR	TEKK	LDV	LQVDD-RR	NAE	1859	59		

Search completed: August 28, 2002, 16:29:02
Job time: 310 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:22:27 ; Search time 29.58 Seconds
(without alignments)
3794.197 Million cell updates/sec

Title: US-09-762-311-5

Perfect score: 6055

Sequence: 1 MEPIFTARKHLLPNEVSD.....RSAEPSDREPTQPEPTGD 1168

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4976	82.2	1141	2 T29104	Tbcl protein - mou
2	2828.5	46.7	1299	2 T00261	hypothetical prote
3	1462.5	24.2	1379	2 T13718	pollux gene protei
4	580	9.6	1030	2 T13163	Rab6 GTPase activa
5	523.5	8.6	1142	2 G89453	protein F35H12.2 (
6	465	7.7	894	2 S61015	hypothetical prote
7	421	7.0	876	2 T49801	hypothetical prote
8	397	6.6	480	2 T29370	hypothetical prote
9	387	6.4	1031	2 T38411	probable GTPase ac
10	386	6.4	356	2 T51376	plant adhesion mol
11	343	5.7	772	2 S62481	hypothetical prote
12	317.5	5.2	720	2 S50920	hypothetical prote
13	317	5.2	814	2 T47841	hypothetical prote
14	293	4.8	519	2 T16712	hypothetical prote
15	290	4.8	468	2 T48686	hypothetical prote
16	282	4.7	834	2 T39891	probable integral
17	276	4.6	458	2 G88391	protein R06B10.5 (
18	275.5	4.5	808	2 T25748	hypothetical prote
19	274	4.5	950	2 S64405	MIC1 protein - yea
20	272.5	4.5	419	2 T08683	hypothetical prote
21	270	4.5	1244	2 T19615	hypothetical prote
22	258.5	4.3	1270	2 T26720	hypothetical prote
23	252.5	4.2	528	2 S51887	hypothetical prote
24	247.5	4.1	492	2 S51885	probable membrane
25	247.5	4.1	633	2 S60408	probable membrane
26	243.5	4.0	487	4 S51886	hypothetical prote
27	234	3.9	637	2 S66953	hypothetical prote
28	232	3.8	856	2 T34491	hypothetical prote
29	227	3.7	1938	2 JC5421	smooth muscle myos

30	227	3.7	1972	2 JC5420	smooth muscle myos
31	225.5	3.7	720	2 T40900	micl homolog - fis
32	224.5	3.7	635	2 T41367	hypothetical prote
33	216	3.6	414	2 T39691	conserved hypotet
34	215.5	3.6	544	2 H84790	hypothetical prote
35	212.5	3.5	1979	1 S03166	myosin heavy chain
36	210.5	3.5	1837	2 T41023	probable nuclear p
37	207	3.4	2663	1 S28261	centromere protein
38	202	3.3	1972	1 A41604	myosin heavy chain
39	202	3.3	1992	2 A47297	myosin heavy chain
40	202	3.3	3187	2 JC5837	364K Golgi complex
41	199	3.3	2007	1 B43402	myosin heavy chain
42	197	3.3	376	2 S57867	oncogene 1 - human
43	197	3.3	786	2 S22155	oncogene 1 (tre-2
44	196.5	3.2	1093	2 A47212	transcription fact
45	195	3.2	1976	2 A59252	myosin heavy chain

ALIGNMENTS

RESULT 1

T29104

Tbcl protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T29104

R:Richardson, P.M.; Zon, L.I.

Oncogene 11, 1139-1148, 1995

A:Title: Molecular cloning of a cDNA with a novel domain present in the tre-2 oncogen

A:Reference number: Z20569; MUID:96032578

A:Accession: T29104

A:Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1141 <RIC>

A:Cross-references: EMBL:U33005; NID:g988220; PID:g988221; PIDN:AAA85223.1

C:Genetics:

A:Gene: tbcl

Query Match 82.2%; Score 4976; DB 2; Length 1141;
 Best Local Similarity 87.9%; Pred. No. 9.2e-270;
 Matches 970; Conservative 47; Mismatches 74; Indels 12; Gaps 5;

QY	37	MPMLPWVVAEVRRLSRGSTRKEPTKQVRLCVSPSGRLCEPEPEGRSQOQDPLIYSSIFEC	96
DB	1	MPMLPWVVAEVRRLSGGCKEPTKQVRLWVSPSGRLCEPEPEGRSQOQDPLIYSSIFEC	60
QY	97	KPQKVHKLHNSHDPVSFACLIKEDAVHROSICVFRKADDTQKVPEIISIRQAGKIARQ	156
DB	61	KPQKVHKLHNSHDPVSFACLIKEDAAHRSQSLCVFRKADDTQKVPEIISIRQAGKIARQ	120
QY	157	EELHCPSEFDDTFSKKEVFLFCGRVTVAHKKAPPALIDECIEKFNHVSGRSGSPRPNP	216
DB	121	EELHCPSEFDDTFAKKEVFLFCGRVTVAHKKAPPALIDECIEKFNHVSGRSGRTD-----	174
QY	217	PHRAPGSG-DEPVRRPMRKSFQGLRSLAFRLQDGLRSGSGFFSFSPESDIEHLIS	275
DB	175	-WEAPTQPSAGPGRPMRKSFQGLRSLAFRLQDGLRSGSGFFSFSPESDIEHLIS	231
QY	276	GHNVQPTDTEENRMTLTIGQSEVYLISPDTKKIALEKNFKEISFCQSGIRHVDHFGFI	335
DB	232	GHNVQPTDTEENRMTLTIGQSEVYLISPDTKKIALEKNFKEISFCQSGIRHVDHFGFI	291
QY	336	CRESSGG--GGFHVCVYFQCTNEALVDEIMMTLKQAFVAAVQQTAKAPALCEGCGPLQ	393
DB	292	CRECSGGGGGFHVCVYFQCTNEALVDEIMMTLKQAFVAAVQQTAKAPALCEGCGPLQ	351
QY	394	SLHKLCEIRIGMNSKTKLELQKHLTLTNOEQATIFEVQKLPRLPRNEORENELIISFLR	453
DB	352	GLHKLCEIRIGMNSKTKLELQKHLTLTNOEQATIFEVQKLPRLPRNEORENELIISFLR	411
QY	454	CLYEKKQKEHIHIGEMKQTSQMAAENIGSELPPSATRFLDMLKNAKAKSLTESLESILS	513

Db 412 CLYEKQEHSTGAPKQTLQVAENIGSDLPSPSRFLDSLKNRKRSLTESLSLS 471
 QY 514 RGNKARGLOEHSISVDLSSLSNTSKPSVCEKALPISSESKLLGSSDLSDDS 573
 Db 472 RGNKARGLOEHSISVDLSSLSNTSKPSVCEKALPISSESKLLGSSDLSDDS 531
 QY 574 ESHLPEEPAPLSPQAFRRANTLSHFFTECOEPPQAPARGSPGVQKLMRYHVSSTETP 633
 Db 532 EGHIAEESALUSPQAFRRANTLSHFFVECPAPPEAOSPGVQKLMRYHVSSTETP 591
 QY 634 HERKDFESKANHLGSGGTPVKTTRHRSWRQIIFLAVATPQKACDSSRYEDYSELGELPP 693
 Db 592 HERKDFESKANHLGSDTCTPVKTTRHRSWRQIIFLAVATPQKACDSSRYEDYSELGELPP 651
 QY 694 RSLPEVCEDEGPPPEEKRTSRELWQKALIQIILLRMKEKQKQASNDLLN 753
 Db 652 RSLPEVCEDEGPPGYRKGRGRHASFESCGKRPSCSRSLVRMEKQKQASNDLLN 711
 QY 754 KRLKLDYEEITPCLKEVTVVWEKMLSTPGRSKIKEDMEKMSAVGQVPRHHRGEIWKFL 813
 Db 712 KRLKLDYEEITPCLKEVTVVWEKMLSTPGRSKIKEDMEKMSAVGQVPRHHRGEIWKFL 771
 QY 814 ABQFHLKHQFPKQKQDPVYKELKQLTSQOHAILEDLGRFTTPHYPFSAQAGQLSL 873
 Db 772 ABQFHLKHQFPKQKQDPVYKELKQLTSQOHAILEDLGRFTTPHYPFSAQAGQLSL 831
 QY 874 YNLKAYSLLDOEVGYCGGLSVAGILLHMSSEAFKMLFMDGMLRQYRPMIIL 933
 Db 832 YNLKAYSLLDOEVGYCGGLSVAGILLHMSSEAFKMLFMDGMLRQYRPMIIL 891
 QY 934 QIOMQSLRLHLDYHRLDLYNHLDEHEIGPSLYAAPFMTFASQPLGFVARVDMIFLQ 993
 Db 892 QIOMQSLRLHLDYHRLDLYNHLDEHEIGPSLYAAPFMTFASQPLGFVARVDMIFLQ 951
 QY 994 GTEVIFKVALSLGSHKFLILOHLENLETIVDFIKSTPLNGLVQMEKTINQVFMEDIAK 1053
 Db 952 GTEVIFKVALSLGSHKFLILOHLENLETIVDFIKSTPLNGLVQMEKTINQVFMEDIAK 1011
 QY 1054 LOAYEVEHVHQLQELIDSSPLSDNORMDKLEKTSNLSKQNLIDLEQLQVANGROSLEA 1113
 Db 1012 LOAYEVEHVHQLQELIDSSPLSDNORMDKLEKTSNLSKQNLIDLEQLQVANGROSLEA 1071
 QY 1114 TIEKLSSSKLQKAMLTLEL 1136
 Db 1072 TVEKLLTSESKLQKALTLEVER 1094
 RESULT 2
 T00261
 hypothetical protein KIAA0603 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
 C:Accession: T00261
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
 DNA Res. 5, 31-39, 1998
 A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
 A:Reference number: Z14086; MUID:98290545
 A:Accession: T00261
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1299 <NAG>
 A:Cross-references: EMBL:AB011175; NID:g3043729; PIDN:BAA25529.1; PID:g3043730
 A:Experimental source: brain
 C:Genetics:
 A>Note: KIAA0603
 Query Match 46.7%; Score 2828.5; DB 2; Length 1299;
 Best Local Similarity 48.3%; Pred. No. 8.1e-150;
 Matches 621; Conservative 184; Mismatches 291; Indels 191; Gaps 28;
 QY 21 FGLQIVGLSPVHSLTMTMPLPWVAEVRRLSRQSTR---KEPVTKVRLCVSPSLRCEP 77

Db 36 FRLWVGSGSLDHRITLPLMPLMAEIRRSQKPPAGCGGAPAAAREVILVLSAPFLRCVP 95
 QY 78 EPG-----RSQWMDLIYSSIFECPQQRVHKLIIHNSHDPSPYACLIK--EDAVHRQ 126
 Db 96 APGAGAGSGTSPSATQPNPAVF--IFEHKAQHSIRFIHNSHDLTYFAYLIKAQPDPEQ 153
 QY 127 SICVYFKADDQTKVPEIILSSIRQAGKIARQELHCPSEFDOTF--SKKFEVLFCGRVTV 184
 Db 154 MACHVFRATDSQVDPVIVSSIRQLSKAAWKEDAKSKONEDAFYNSQKFEVLYCQKVT 213
 QY 185 HKKAPALIDECIEKFN-----HVSG--SRGSE-----SPRENPPHAA-- 220
 Db 214 HKKAPSSLLDDCMKFSLHEQRLKIQEQRGPDGDLADLEVVPVPGSGDCLPEADG 273
 QY 221 -----PTGSOEPVRRPMKSFSPQGLRSLAFRELQDGLRSLSGFPSS----- 263
 Db 274 TDTHLGLPAGASQAPALTSSRVCFPE-----RILEDSGDEQOEFRSRCSSVTGVOR 324
 QY 264 -----FEESDIENHLISGHNIIVQTDIBENRTMLFTIQGSEVYLISPDTKKIALEKNFE 318
 Db 325 RVHEGSKSQPRRRHASAPSHVQPSDSEKNRTMLFQVGRFEINLISPDTKSVVLEKNFKD 384
 QY 319 ISFCSQGIHVDHFGFICRESSGGGFHFVYVFCFTNEALVDEIMMTLKQAFVAAVOQ 378
 Db 385 ISSCSQGIKVDHFGFICRESSPEGLSQYICVYVFCASESLVDEVMTLKQAFSTAAALQ 444
 QY 379 TAKAPALCEGCPLOSLLKLCERIEGMNSSTKLEKQLHLTLTNOEATIFEVQKLRP 438
 Db 445 SAKTOIKLCEACPMHSLKLCERIEGLYPPRAKLVQIRHLSLTNOEADIFERVQKMP 504
 QY 439 RNEQRENELIISFLCLYEKQEHIGEMKQTSQMAENIGSELPSPSAT---RFRIDM 495
 Db 505 VSDQENELVILHLQKCEAKQKTHVHIGEGPSTIS-----NSTIPENATSSGRFKLDI 558
 QY 496 LKNKAKRSLTESLSILSRG--NKARGLOEHSISVDLSSLSLTSLNTSKPSVCEKALP 554
 Db 559 LKNKAKRSLTESLSILSRG--NKARGLOEHSISVDLSSLSLTSLNTSKPSVCEKALP 606
 QY 555 ISESEFKLLGSS-----EDLSSDSESHLPEAPLSPQQAQFRRRANTLSHPP----- 602
 Db 607 -----GSPGPTPPASPSSSAWQTFPEEDSD--SQQ--FRRRAHTSHPPSSTKRKL 654
 QY 603 ECQEPPOPARSPGVSQ-----RKLMYRHVST-----ANHLGSGGTTPVTRHRSRQ 663
 Db 655 NLQDGRAQVRSPLLRQSSEQCSNLSSVRRMYKESNSSSLSPSLTSSFSAPSFTAPSL 714
 QY 631 -----ETPHERKDFESK-----RKLMYRHVST-----ANHLGSGGTTPVTRHRSRQ 663
 Db 715 KSFYQNSGRSLSPQYENIRQDTASESDGEGKRKRTSTCSNESLSVGGTSVTPRRLSMRQ 774
 QY 664 QIFLRVATPQKACDSSRYEDYSELGELPPRSPLEPVCEDGPF-----GPPPEEKRR 715
 Db 775 RIFLRVAFPMKSPSAMQOQDGLDRNELLPLSPLTMEEPVIFLSEDDPEKTEERK 834
 QY 716 TSRELRLEWQKAILQOILLRMKEKQKLO--ASENDLNRKLUKLYEETPCPKVETVW 774
 Db 835 KSKELRSLWRKAIHQOILLRMKEKQKLO--ASENDLNRKLUKLYEETPCPKVETVW 894
 QY 775 EKMLSTPGRSKIKEDMEKMSAVGQVPRHHRGEITWKFLAEQFHLKHQFPSPKQKQDPVY 834
 Db 895 DKKLLN--CRAKIRCDNEDITHLLKGVGPKSRRTGTWQFALQYRLHRLPNKQPPDISY 953
 QY 835 KELLQKLSQHAILEDLGRFTTPHYPFSAQAGQLSLYNILKAYSLLDQEVGYCQGLS 894
 Db 954 KELLQKLSQHAILEDLGRFTTPHYPFSAQAGQLSLYNILKAYSLLDQEVGYCQGLS 1013
 QY 895 FVAGILLHMSSEAFKMLFMDGMLRQYRPMIILQIOMYQSLRLLHLDYHRLDLYN 954
 Db 1014 FVAGILLHMSSEAFKMLFMDGMLRQYRPMIILQIOMYQSLRLLHLDYHRLDLYN 1073
 QY 955 LEEHIGISLYAAPFLMFAQSQPLGFVARVDMIFLQGTVEIFKVALSLGSHKFLIL 1014

QY 812 FLAEPFLKHQFSPKQOPKDVYKEL--LKQLTQQHAILIDLG-RTFPTHVPFSAQLGA 868
Db 433 VLQO-----SKNELETIVRDLDEKWEMLRLEKVIIRRDGART--SYSFRAAAGL 481
QY 869 GQLSLYNILKAYSLLDQEVGYCGGLSFVAGILLHSHSEEEAFKMLKFLPMFDMGLRKOYRP 928
Db 482 -QEGFGVCYKAYALFEAVGAYQGMFLVMPPLFNPMPEEAFCLLVRLMNYHLRELFIQ 540
QY 929 DMILIQMTQLSRLHLDYHRLDLYNHLHEEIGPSLYAAPFWLTMFASQPLFLGVARVFD 988
Db 541 DMPGLHKCLYQFERILEDELPALYCHLHRRGISPHLYATQWFLTLFAYRPPQLVLRIDY 600
QY 989 MIFLQGTGVFKVALSLGSHKPLIHOHENLETIVDFIKSTPLNGLVQMEKTIQNOVFEM 1048
Db 601 LIFEGLSALIKGIVLMKNATALLGMSDMSQLTTFILKDRLEFDV-YIDATPSSNILEN 659
QY 1049 DIAQQAQYEVHVHQLSEELIDSSPLSDNORMDKL-----EKNSSLRKQNLDLLEQ 1100
Db 660 GFFGSSA-SIDKEVYRADQLVRDADCVNITAEALLKAYGKEWEEKTKA--EKEREQBLEG 716
QY 1101 LOVANGR---IOSLEATIEKLSSSKLQKQAMLTLELERSALLQIVVEELRRSAE 1152
Db 717 LKQANTNYTVVRRLEERIEAVDREQASLATELVRTKVENELRDENESLRGQVKE 772

RESULT 8

T29570

hypoetical protein T24D11.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29570

R:Wu, X.; Gattung, S.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid T24D11.

A:Reference number: Z20643

A:Accession: T29570

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-480 <WUX>

A:Cross-references: EMBL:U49940; PIDN:AAA93411.1; CESP:T24D11.1

C:Genetics:

A:Gene: CESP:T24D11.1

A:Introns: 59/1; 91/2; 123/3; 192/3

Query Match 5.6%; Score 397; DB 2; Length 480;
Best Local Similarity 25.7%; Pred. No. 9.9e-15;
Matches 123; Conservative 83; Mismatches 158; Indels 114; Gaps 15;

QY 661 WRQOI-----FLRVATPQKADSSSRRYEDYS---ELGELPPRSPLEPVEDGPGFP 708
Db 26 WKRLLETSPPHFFPAPATTTTACIFGSKLVAGSADTECGE-----PCETG---- 72
QY 709 PPEKRTSRELRLMOKAYLQOI-LLLRMKEKNQKLOASENDLLNKRLL----- 756
Db 73 -----APVSLNEVDLLAKMEQLNKSNEEDSRVASKTSSSRKRGARE 116
QY 757 ---KLYEETPCLEKVTTWKMLSTPGRSKTKFDMKMHSAVGQVPHRHGETWKFL 813
Db 117 HSPDEEDLWSYNGELILNWE-----IEVKRPNYIKDLVKRGIPQHFRMAWQNL 168
QY 814 AEQFHLKHQFSPKQOPKDVYKELKQLTQQHAILIDLGRTPTTHPYFSAQLGAGQLS 873
Db 169 SN-----ASVSSVHDL-YSDMKQSSVYEVKIQIDIPRTYPELDFFK-DGERQSLL 218
QY 874 YNLIKAYSLDQEVGYCGGLSFVAGILLHSHSEEEAFKMLKFLMFDMLGRKOYRPMIIL 933
Db 219 FNVIKAYSHVDKEVGYCGGSAFTVGLLLQMPBEEAFVLVSLMENVRLRYLKYKPTMTDL 278
QY 934 QIQMYQLSRLHLDYHRLDLYNHLHEEIGPSLYAAPFWLTMFASQPLFLGVARVFDIFLQ 993
Db 279 GLCMFQLECLVQDQMPDLXTFFHNNMGFDTSKYASSWFLTLFTTTMTPLDIANRIMDCFLVE 338

QY 994 GTEVIEFKVALSLGSHKPLIHOHENLETIVDFIKSTPLNGLVQ-MEKTINQOVFEMD--- 1049
Db 339 GMDFFICISIA-----ILQQARIELL-----RLDMGMLKYFOREVREYEFADL 384
QY 1050 ---IAKQLQAQYEVHVHQLSEELIDSSPLSDNORMDKLEKTNSSLRKQNLDLLEQLQVA 1104
Db 385 LFTVANQVQL-----NAKRMKLEKDYLTFRKTKQEQAVALRVS 423

RESULT 9
T38411
probable GTPase activator protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38411
R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21792
A:Accession: T38411
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1031 <BRO>
A:Cross-references: EMBL:Z73100; PIDN:CAA97366.1; GSPDB:GN00066; SPDB:SPAC26F1.09
A:Experimental source: strain 972h; cosmid c36F1
C:Genetics:
A:Gene: SPDB:SPAC26F1.09
A:Map position: 1

Query Match 6.4%; Score 387; DB 2; Length 1031;
Best Local Similarity 20.7%; Pred. No. 1.1e-13;
Matches 207; Conservative 153; Mismatches 377; Indels 262; Gaps 39;

QY 206 SRGSESPRNPHPAAPTGSOEVPVRPMRKSFQPGI-RSLAFKELQDGLRSGSGFFSF 264
Db 144 SESTEFPVNEVNETATLGNEDVGE--RSGFPEGLDNEPESQRLDETG-----NL 192
QY 265 EESDIENHLISGHNIVQPTDIEENRTMLFTIGQSVLYLISPDTKKIALE-----KNFKEI 319
Db 193 APEDLAKDEVKSVHEFNEPNDLRQ-----QEEYSDDDDINVEFEVDVNIENHQL 243
QY 320 SFC--SQGIRHV-DHFGFTICRESSGGGFHFVYVQCTNEALVDEIMTMLKQAFVAAV 376
Db 244 SVADEQTSRLVKGMIFVGKEDFGEA-----DISNSVFEIQ---NGPNSDIVSGF 292
QY 377 QQTAKAPALCECPQLQSLHKLKEREGMNSSKTKLEKHLTLTNOEQATIFEVQKL 436
Db 293 KETSSI-----VNSSST-----TEKPGVALDSQNDTSIF----- 321
QY 437 RPRNEQRENELIISFLRCLYEEKQKHIIGEMKQTSQMAAENIGSELPPSATFRRLDML 496
Db 322 ---NEQTSNLSLTF-----NDLTLDHL-----PENVESE--PVACK----- 353
QY 497 KNAKAKSLTESLESILSRGNKARGLQHSISVDLSLSLSTLSTNTSKPSVCEKE----- 551
Db 354 -----ENETAKNESGASDNDHKANVHFVLKSSDAITLNEEKIATQDDPLEA 401
QY 552 ALPISSESKLLGSSSEDLSSDSSEHLPPEAPLSPQAPRRRANTLSHPPIEC-----OE 606
Db 402 PTPIVASSSTIFUNSNQRNDELASGSQEPHP-----KGTNSTSSSLPLDNTNNLSNE 454
QY 607 PPQAPAGSPGVSRQKMLRVHSVSTETPHERKDFESKANHLGDSGGTP-----VKRRHSW 661
Db 455 PPSHVLDASSETIEVQTIKKLQNVQPETIKDEVGKN-TAFSPGTSLSNTHVTKSRSA 513
QY 662 ROQIFURVATPQKACSSSRKYEDYSELGELPPR---SPLEPVCEDEGPFPPPEEKRTSR 718
Db 514 HNNSTSPFSTAVSWLNPLRYP-----SOKSPRVISYLESVFISKP-----R 556
QY 719 ELRELQKAILQOILLRLMEKENQKLOASENDLLNKRLLKLDYEITPCLKEV----- 770
Db 557 SIGDAOKLEILEYLOSQSSSTVSNQVFTLLSNFTLQNPLFVLJD-----ECDFEFLNLMHN 611

QY 771 -----TTVWEKMLSTPGRSKIKFDMKMH-----AVGQVPRHRHGRGIWFLA 814
Db 612 SHTVHTVWKTI-----SSWTSYDYEMQYSLSKNCSDSKAIRKLDRTTPEI-----661
QY 815 EQFLKHQFPSPKQ--PKDVPYKELLKQLTSQQHAILDLIGRTPTTPPYFSAQAGAGLS 872
Db 662 ----LSHFFSNRQLEPTD-----NIAESTAN 684
QY 873 LYNILKAYSLLDQEVGYCQGLSFVAGILLHMSSEAFKMLKFLMDFMGLRKQYRPMII 932
Db 685 LHRVLSLAIVLPQVGTQGSWIGALLMLHPAPQAFALLVFLKNYHLQNIFFSEMRG 744
QY 933 LQIQWQLSRLLHDYHRLDYNHLEHEIGPSLYAAPFLTMFAQFVARFDMIFL 992
Db 745 LSRVLHQFTLVEDYMPSLATHFRRQDIKTCYASEMFLTLFAYKFLPVVAHYLDLFL 804
QY 993 QGTEVIFKVALSHKPLJLOHENLTIIVDFIKTLPLNGLVQMEKT-----1041
Db 805 YPGILFNGALL-SHSQESLLKLMMDRLISYLKEDI----FLAFKETQGENYDTSLF 859
QY 1042 INQVEMDIKQ--IQAYEVEYHVL--QBELIDSSPLSDNOMDKLEKTNSSLRKONL---1095
Db 860 VKTAFSEFIQPDVLDRYGNEYDILLKSEHLDSS-----LEEMRNHKSLSNEHFIMLS 912
QY 1096 DLLQQLQVANGRIOSLEATIEKLSSSESKLQAMLTLEL 1134
Db 913 DSMANLQVHEHNMGSAL-LLKERMVLYKNOTVQASLSKSEI 950

RESULT 10

T51376

plant adhesion molecule 1 (PAM1) - Arabidopsis thaliana

N:Alternate names: protein FIN13_70

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Nov-2000

C:Accession: T51376; T52031

R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Ban

submitted to the Protein Sequence Database, August 2000

A:Reference number: 225393

A:Accession: T51376

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-356 <SAT>

A:Cross-references: EMBL:AL391145

A:Experimental source: cultivar Columbia; BAC clone FIN13

R:Stevenson, B.A.; Zhu, J.K.

submitted to the EMBL Data Library, June 1998

A:Reference number: 225910

A:Accession: T52031

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-356 <STE>

A:Cross-references: EMBL:AF069528; PIDN:AAC33763.1

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 5

A:Introns: 48/2; 113/3; 170/3; 209/3; 275/3; 328/3

A:Note: FIN13_70

Query Match 6.48; Score 386; DB 2; Length 356;

Best Local Similarity 31.4%; Pred. No. 2.6e-14;

Matches 85; Conservative 63; Mismatches 107; Indels 16; Gaps 4;

QY 742 OKQASNDLLNKLKLDYEITPCLEKVTWVEKMLSTPGRSKIKFDMKMH---SAVG 798

Db 28 QEHSSSPORFTKSSINYEK---EKRVTYKWKRMIGTGGSDWKHYVRPVPVVKRRIR 83

QY 799 QGVPRHRRGIWFLAEQFLKHQFPSPKQKDPYKELLKQLTSQQHAILDLIGRTPT 858

Db 84 KGIPDCLRGLWQLISGRDLLANPG-----YVQLVIYVETSASELDIIRDSIRTPS 137

QY 859 HPYFSAQAGAGQLSLYNILKAYSLLDQEVGYCQGLSFVAGILLHMSSEAFKMLKFLM- 917
Db 138 HVFFQKRHGPGQGRSLYNILKAYSYVDKRGVYQVGMGTAGLLLYMSEDAFWLLVALLK 197
QY 918 --PDMGLRKQYRPMIILQIQWQLSRLLHDYHRLDYNHLEHEIGPSLYAAPFLTMFA 975
Db 198 GAVHSPTEGLYQAGLPLVQVQLLQFDQLVRELMPKLGEHFTQEMINFSWYASQWFIIVFS 257
QY 976 SQFPLGFVARFDMIFLQGTVEVIFKVALSLL 1006
Db 258 YSLFPHSALRIWDVFLAEGVKIVFKVGLALL 288

RESULT 11

S62481

hypothetical protein SPAC4G8.04 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 31-Jan-2000

C:Accession: T38849; S62481

R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, October 1995

A:Reference number: Z21745

A:Accession: T38849

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-772 <BA2>

A:Cross-references: EMBL:Z56276; NID:g1022345; PIDN:CAA91205.1; PID:g1022349; GSPDB:g

A:Experimental source: strain 972h-; cosmid c4G8

C:Genetics:

A:Gene: SPAC4G8.04

A:Map position: 1L

A:Introns: 680/3

Query Match 5.7%; Score 343; DB 2; Length 772;

Best Local Similarity 22.1%; Pred. No. 2e-11;

Matches 198; Conservative 126; Mismatches 340; Indels 232; Gaps 34;

QY 294 TIGQSEVYLISPDTKKIALEKNFKFISFCSGIRHVDHFGICRESSGGGGFHFVCYVFO 353

Db 4 SVSDSSVYLSAPTAKPIA-----SGSVATG-----28

QY 354 CTNEALYDEIMMTLKQAFVAAVOOTAKAPALCGCCPLOSILKLCERIEGMNSKTKLE 413

Db 29 -----IDS--MPSKVDITPCDLLNKSAPLFVECNQESLHSPGSLHVPASIERL 80

QY 414 LQKH-LTTLTNQEQATIFE---VQKLPRNEQRENELIISFLRCLYEEKQEH---IHI 466

Db 81 IEKHGAVNLLRQLAKDVAERDSFISDLKFHFESE-----YVFRELLREHGLDPLV 131

QY 467 GEMQTSQMAAENIGSELPPSATR-----FRDLMLNK---AKRSILTESLE 509

Db 132 ANTKLSQRHSASFPPSSQEPSIPENPSLTGKPHLYARIDSAINEPTPSDRSLSPSLV 191

QY 510 SILSRGNKARGLOEHSISVDLSDSLSTLSNTSKPSYCEK-EALPISE-----SS 559

Db 192 PLL---KLPAI-DHAYSSSSSDLPSPDPSASVYASSKQKASSLKLTSLKKFYWS 246

QY 560 FKLGGSEDLSSDESHLPEEPAPLSPQAFRRANTLSHPPIECQBPQPARGSPGVQ 619

Db 247 SSLQHTRENLDSTSLRDHDPDLLSSSKPPRSSP-----RCSTPSVSTFVSATSE 298

QY 620 RKLMYHSVSTETPHERKDFESKANHLGDSG-----651

Db 299 PEVETY-SVSTKNSSSNKLRSLSKLLSTNLNKNKPLSLSTAPSMPSIGSVELGNMIP 357

QY 652 --TPVKTRRSRQOIFLRVATPOKACDSSSYEDYSELGELPPR-SPLEPVCEGDPGP 708

Db 358 KETOPPSNRNDWKDYL-----DNNSK-EILOQFGFLQKRPSTDTPLC-----398

QY 709 PPPEB-----KKRT-----SRELRLWQKAIL---QOILLRMEKNQKLQA 746

Db 709 PPPEB-----KKRT-----SRELRLWQKAIL---QOILLRMEKNQKLQA 746

Db 399 -PEDIKLNQKOTLSFYESNYGLVDDFGNELDGLNDSPILLNKKDILLDMKESTRQKNWS 457
QY 747 SENDLLNKKLDYBEITPCLEKVTWVWVKMLSTGRSKIKFDMEKMHSAVQGVPRHR 806
Db 458 LFFORLYKKYKT-DEDTIGLIGISSGVK-----GRHG-KRWKHKFRELKNGVPLCYK 510
QY 807 GEIWKFLABQFHLKHOFPSKOQKDPV--YKELLKQTSQQA-----ILIDLGRTFPTHP 860
Db 511 AKVWLECSGAYQL-HS-----PGYEEILLSRTDEVSASVAQIDMDINRTMAKNV 559
QY 861 YPSAQLAGQOLSYNLKAYSLDQEVGYCQGLSFV-AGILLHMSSEAFKMLFMPD 919
Db 560 FEGGK-GPGIPKRLRLVAYSRRHNPHIGYCOGMNIGAFLLLLYASEDAFYMLSIEN 618
QY 920 MGLRQYRDMILQIOMYQLSRLLDYHRLYHLEHEIGPSLYAAPWFLTMSASQPP 979
Db 619 VLPPKYFTPLMTSRADQLVLSFVKESLPEIYSHLELLGVDLDAISFWFLSVTDTLP 678
QY 980 LGFVARVDMIFLOQTEVIFKVALSLGSHKPLIIOHENLETIVDFIKSTPLNLGLVOME 1039
Db 679 TNISFRIFDMLFCGVCVFLRVALATILSKQIILACNSSSAYISFLSDL----- 728
QY 1040 KTNQVFMFMDIAKQLOAYEVEHVLOEBELIDSSPLSDNORMDKLEKTNSLRKQNL 1095
Db 729 -----VOYSFQDPSFIKEAA-----DRWSKLVTEKSIERKNL 761

RESULT 12

S50920
hypothetical protein YMR192w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR192w (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S50920
R:Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A:Reference number: S50917
A:Accession: S50920
A:Molecule type: DNA
A:Residues: 1-720 <PEA>
A:Cross-references: EMBL:247815; NID:g642280; PIDN:CAA87813.1; PID:g642284; MIPS:YMR192w
C:Genetics:
A:Map position: 13R

Query Match 5.2%; Score 317.5; DB 2; Length 720;
Best Local Similarity 21.0%; Pred. No. 4.9e-10;
Matches 149; Conservative 126; Mismatches 24; Indels 191; Gaps 26;

QY 557 ESSFKLLGSSEDLSS--DSSEHLPEE-----PAPLSPOQAFRRRANTL--SHFTPECOE 606
Db 92 ESQSKTILPSDDLSQLQLETEESKVEEALKRITSPPLP-----RADCTEESASAKSSL 145
QY 607 PP-----QPARGSPGVSKRLMYHVSVTETPHERKDFESKANHLGDSGGTPVKRR 658
Db 146 PPVLGNKNDQAPLDRPOLPQPVVNAETLHLKAPGN----- 183
QY 659 HSRQOIFLRVATPKACDSSRYEDYSELGLPRP--SPLEPCVEDGPFPPPEKKR 715
Db 184 -----ATPSKSPTSVAGNSSSTPTPLPPRIEDLDLAAQ----- 219
QY 716 TSRELRLQKAILQIOLLRLMEKENQKLOASENDLLNRLKLDYEEI--TPCLREVTTV 773
Db 220 -----KHFLASTFKRNLKLYKSEDSIKCDLKN-ILN--LKEDSKINNNEIPEVSVF 271
QY 774 WEKMLSTPGRSIKTFDMEKMHSAVQGVPRHRGETWKEFLAQFHLKHOFPSKOQKDPV 833
Db 272 WLKVIQDYQNILIN-DIETLHLQSLRGIPAAVRLVWQLVSY-----AKRSFDFI 321
QY 834 YKELLKQTSQQAIIIDLGRTFPTHPYFSAQLAGQOLSIX-----NILKAYS 881
Db 322 YETYLTEMAP-----FDVQEFNQLKMDVEVPSEYVKRISNVLKAYL 363

QY 882 LLDQEVGYCQGLSFVAGILLHMSSEE--AFKMLKFLMDFMGLRKOYRDPDMLIQLQIMVQ 939
Db 364 LFDPECFESTDIAYIIN-MILDVCEENAFGLLRLKVKYGLRLFLPASEIDILCYK 422
QY 940 LSRLLDYHRLYHLEHEIGPSLYAAPWFLTMSASQPPFLGFVARVDMIFLOQTEVIF 999
Db 423 FDRLEVEEYFPEIHNHMKVGRSSMPLPGFFTTLFOKKLPTETIQPRIGMDVFEGLDSIM 482
QY 1000 KVALSILGSHKPLIIOHENLETIVDFIKSTPLNLGLVOMEKT-----INQVF 1046
Db 483 RILATLLSNSRDHLK-WGFDDMLLEKSLGLLDAYIKQNDGTRGDTLLSNECDMLLQDS 541
QY 1047 EMDIA---KQLOAYEVEYHVL-----QBELIDSSPLSDNORMDKLEKTNSLRKQNL 1094
Db 542 MMKVAITPKTKMKYSEVEIHRLDNEKEVQYKSTETKLNHLQKVKRKLNDYTSLSRHH 601
QY 1095 LDLLLQ-----LQVANG-RIQSLEATIEKLLSESK-----LQK 1127
Db 602 VTIANELVKNRNLNIEVNLNENNGYKLIID--LKKLQSEKKKQVLGVVYPNDLKKDLLE 659
QY 1128 AM-----LTLELERSALLQVTEELRRRSABPSDREPECTOPEPTG 1167
Db 660 TMKKNTQVMDENKLDQRISELERLIEIKTANKNGTLFEYSNKNPLG 709

RESULT 13

T47641
hypothetical protein T15C9.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47641
R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224470
A:Accession: T47641
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-814 <MEW>
A:Cross-references: EMBL:AL132970
A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetics:
A:Map position: 3
A:Introns: 28/2; 51/3; 252/3; 312/3; 350/3; 370/2; 392/3; 410/1; 441/1; 469/1; 604/3;
A:Note: T15C9.20

Query Match 5.2%; Score 317; DB 2; Length 814;
Best Local Similarity 22.6%; Pred. No. 6.2e-10;
Matches 151; Conservative 108; Mismatches 220; Indels 188; Gaps 28;

QY 645 HLGDSGGTPVKTRR-HSNRQOIFLRVATPKACDS--SSRYEDYSELGELPPSPLEPVCE 702
Db 26 HKRDAYGPPVRPQHVRQYREYADIVKEEERSDRWSSFLDHWESTELPTNGSSSENI-- 83
QY 703 DGPFGPPPPPEKRTSREL------ELAQ 725
Db 84 HAPFS---ESEKEKELNKGFGEDLHTDKGSDVTPDNASEEGHPDAKKNVHVRVOLT 140
QY 726 K-----AILQOILLRLMEKENQKLOASENDLLNRLKLDYEEITPCLEKVTWVWVKMLST 780
Db 141 ETRPSLRSTEDLMSIRVKKGD-LSKSEQ-----APVK-ISPFDACKS 185
QY 781 PGRSKIK-----FDMEK-----MHSVAGQGV 801
Db 186 KCADSDSEDFYDVERSVDQGSSTSDGTGVSIGIPVAADASPLSTCPMKVEEVLIRGCV 245
QY 802 PRHHRGEIWK-FLAQFHLKHOFPSKOQKDPYKELK-----QLTSQHA----- 847
Db 246 PMALRGELWQAEVGR-----KRRKDY-YONLLAADSGSVNTIEQDMQHVDDKGS 295
QY 848 -----ILIDLGRTFPTHPYFSAQLAGQOLSILYNILKAYSLDQEVGYCQGL 893
Db 296 STESAVVEKWKQIEKDLPRTFPGHPALDDD---GRNALRLLLTAYARHNSVGYCQAM 352

Qy 894 SFVAGIILLHMSSEAEAFKMLFMDGLRKQYRDPDMLIQLQMYQLSLLHHDYHRDLN 953
Db 353 NFFAALLLLPEENAFWALIGLIDDY-FNGYISEMIESQVDQLVLEELVRERPKLVH 411
Qy 954 HLEEEHIGPSLYAAPWFTMFASQPLGFVARVFMIFLQTEV-IFKVALSLLGSHKPL 1012
Db 412 HLDYLGQVAVWGTGFWFSLFNNMPLWESVLVWDVLFEGTRVYMLFTALAMELYGPA 471
Qy 1013 ILQHENLETVDFIK-----STLPNGLVOME-----KTINQVEMDI----- 1050
Db 472 LVTTKDAGDAVTLQSLAGSFDDSSQLVLTACMGQYVHEIRLQELRSKRPVAVIALEE 531
Qy 1051 -AKQLQAYEVE-----YHVLQEE-----LIDSSPLSDNQRWDKLEKTSLSRKQNLDLLE 1099
Db 532 RSKGLQARWDSKGLASKYLNFKQDPKSVLQDSKASLSNGSLSRSSEGSNADEVLSVIG 591
Qy 1100 QIQVANGRIQSLAEATIE--KLTSSESKLQKAMLTLELERSALLOVVEILRR--SAEPSD 1155
Db 592 DGEVDS--VQDLQAQAECLKLEEK--RSALLRAEELEIALMEIVKEDNRRQLSAKVEQ 646
Qy 1156 REPECTQ 1162
Db 647 LEQEMAE 653

RESULT 14

T16712

hypothetical protein R1185.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T16712

R:Leinbach, D.

submitted to the EMBL Data Library, December 1995

A:Description: The sequence of C. elegans cosmid R1185.

A:Reference number: 218564

A:Accession: T16712

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-519 <LEI>

A:Cross-references: EMBL:U42831; NID:g1123122; PID:g1123123; PIDN:AAA83513.1; CESP:R1185

C:Genetics:

A:Gene: CESP:R1185.1

A:Introns: 13/2; 48/1; 85/2; 146/3; 190/2; 226/1; 248/3; 381/2; 473/3

Query Match 4.8%; Score 293; DB 2; Length 519;
Best Local Similarity 21.3%; Pred. No. 7.1e-09;
Matches 140; Conservative 103; Mismatches 237; Indels 178; Gaps 25;

Qy 497 KNAKRSLSLESLSRGNKARGLOEHSISVDLSDSSLSLTSNWSKPSVCEKEALPIS 556
Db 8 KNAAGLCLEPLCSL-----SLNTRDCSSSSSEQDYDSGIGCTE-----S 49
Qy 557 ESSFKLLGSSDLSDESHPLEPAPLSPOAFRRRANTLSHRFIECQPPQARGSPG 616
Db 50 DSKRSSLRSSLSDFDEP--PSKETLSKSRFF-----NFP----- 84
Qy 617 VSQRKLRYHVSSTETPIERKDFESKANHLGDSGTPVTKRRHSWRQIFLRVAPPQAK 676
Db 85 -----KNFFSRNKEKGWKFSGNRKQSGVMATGL----- 116
Qy 677 DSSSRYEDYSELGELPPRSPLEPVCEGDFGPPPEKRTSRRELWOKAILQOILLR 736
Db 117 -----ILEERPGLPSKSADEAAHH-----KOMYLDI 143
Qy 737 MEKENQKLOASNDLLN---KRLKLDYEITPC---LKEVTVWEKMLSTPGRSRIKEDM 790
Db 144 LEQAAKKQRAEKERLQAKAEQKRLSEQTAACHRVVWVEQILPKWDEMK-----DS 193
Qy 791 EKMHSVAGQVPRHRHGTWKF-LAEQFHLKHOPPSKQPKDVPYKELLKOLTSQOH--- 846
Db 194 KRCRELWQGVPAKVRGELWFLFTIGNQIEITKELYDGLM--DQAEKIAKQLAEQNKNSA 251

Qy 847 -----AILDLGRTPPT-----HPYFSAQLGAGQLSYLNILKAYSLLDQEVGYCQ 891
Db 252 ERKETSVTQIHLDAFRTFTSLGMFOKDGPPYD-----HLLKLSAYAILRPIGIVYQ 303
Qy 892 GLSFVAGIILLHMSSEAEAFKMLKFL-MFDMGLRKQY-----RPMIILQIQMYQLSRLLDH 946
Db 304 SMTFAAVALLLQMDYPAP--ISFANLLDRSLQSAFFGLKQPMTEYFI---AYDRYLBQ 358
Qy 947 YHRDLYNLEHEHIGPSLYAAPWFTMFASQPLGFVARVFMIFLQTEVIFKVALSLL 1006
Db 359 ELPALHQHLDKLDVAPDLYLIEWTFAMYAKSLPDLVTCRIMDVYFRDGEETFKAAALGIL 418
Qy 1007 GSHKPLIILQHENLETVDFIKSTLPN-LGLVQEMXTINQVEMDIKQALQAYEVEVHVLIQ 1065
Db 419 RMIEPKLITMD-FDQCVFEL-TKLPNTLTGAEFLRNIEPFMRPYNGESRSKRRSFIQ 476
Qy 1066 EELIDSSPLSDNQRWDKLEKTSLSRKQNLDLLEQLQVANGRIQSLAEATIEKLLSSES 1123
Db 477 E-ID-----ERYNPGTGTARTQITHNVQELKMS---KSLSGFIKDLLSSPS 518

RESULT 15

T48686

hypothetical protein DKFZp761D1823.1 - human

C:Species: Homo sapiens (man)

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000

C:Accession: T48686

R:Boecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, April 2000

A:Reference number: 224536

A:Accession: T48686

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-468 <AAA>

A:Cross-references: EMBL:AL353935

A:Experimental source: adult amygdata; clone DKFZp761D1823

C:Genetics:

A:Note: DKFZp761D1823.1

Query Match 4.8%; Score 290; DB 2; Length 468;
Best Local Similarity 24.9%; Pred. No. 8.9e-09;
Matches 121; Conservative 73; Mismatches 181; Indels 110; Gaps 17;

Qy 664 QIFLRVATPQKA-----CDSSRYEDYSELGELPPRSPLEPVCEGDFGPPPE 711
Db 20 KIWRKVAEKERALLTKCAYLQARNQVSKY-----LAGRLRLQALGD----- 63
Qy 712 EKKRTSRELRELWOKAILQIILLRWEXENQKLOASENDLLNKRKL-----DYE-EIT 764
Db 64 EASECSELLRLQVQEA-----LOWEAGEASSDSIELSPISKYDEYGLFTVPDYEVDL 116
Qy 765 PCLKEVTVWEKMLSTPGRSRIKEDMKNHSAVGO-----GVPRHRHGEIWK 811
Db 117 KLLAKIQALESRSHELLGLEAVDRPLRERWALGDLVPSAELKQLLRAGVPREHRPVWR 176
Qy 812 FLAEOFHKLKHOPPSKQPKDVP--YKELLKOLTQSOH-----AILDLGRTPFTTHYFSAQ 865
Db 177 WLW----HL-----RVQHLHTPGCYOELLSRGQAREHPAARQIQLDLNRTFPNNKHFTCP 227
Qy 866 LGAGQLSYLNILKAYSLLDQEVGYCOGLSFVAGIILLHMSSEE-AFKMLKFLWFMQGLRK 924
Db 228 TSSFPDKLRRVLLAFSWQNPTIGYCOGLNRLAAIALLVEEESAFWCLVAIVETIMPAD 287
Qy 925 QYRPMIILQIQMYQLSRLLDYHRDLYNLEHEHIGPSLYAAPWFTMFASQPLGFVA 984
Db 288 YCNLTASQDQVRLQDLSEKPLRLMAHLQGHVHDLSTFTNFWFLVFADSLISNILL 347
Qy 985 RVFDMIFLQTEVIFKVALSLLGSHKPLIQLHEN---LETIVDFIKSTLPNGLVQMEKT 1041
Db 348 RWDADFVEGTQVVPVRYALATFKYNEKEILRLQNGLEIYQYLRFTTKTISN-----SRKL 402

QY 1042 INQVFENDIAKQLQAYEVEYHVLQEELIDSSPLSDNQ-----RMDKLEKTNNSSLRKQNLDL 1097
Db 403 MNTAFN-----DNPFPMKQLROLRWVHRERLEAELRE----- 435

QY 1098 LEOLQ 1102
Db 436 LEOLK 440

Search completed: August 28, 2002, 16:24:25
Job time: 118 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:22:27 ; Search time 38.61 Seconds
(without alignments)
3360.119 Million cell updates/sec

Title: US-09-762-311-5

Perfect score: 6055

Sequence: 1 MPEITFTARKHLLPNEVSD.....RSAEPSDRPECTQPEPTGD 1168

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6055	100.0	1168	AAV45096	Human TBC-1 protei
2	4976	82.2	1141	AAW44777	Human tbc-1 protei
3	4823	79.7	937	AAW38698	Human polypeptide
4	4772	78.8	981	AAW43195	Human ORFX ORF2959
5	3891	64.3	763	AAW40484	Human polypeptide
6	3463	57.2	674	AAW5675	Human protein sequ
7	1600.5	26.4	343	AAU21586	Novel human neopla
8	1559.5	25.8	390	AAU21586	Novel human neopla
9	1467.5	24.2	1379	AAW57823	Drosophila melanog
10	1334	22.0	265	AAW93840	Human polypeptide
11	588.5	9.7	1119	AAW40436	Human polypeptide

12	580	9.6	1069	22	AAW38650	Human polypeptide
13	580	9.6	1069	22	AAW38650	Human polypeptide
14	492	8.1	1194	22	ABW62985	Drosophila melanog
15	481	7.9	94	20	AAW11605	Human 5' EST secre
16	479.5	7.9	803	22	ABW65883	Drosophila melanog
17	392.5	6.5	353	21	AAW606209	Arabidopsis thalia
18	386	6.4	356	21	AAW606210	Arabidopsis thalia
19	382	6.3	292	21	AAW606210	Arabidopsis thalia
20	371.5	6.1	749	22	AAW39630	Human polypeptide
21	371.5	6.1	749	22	AAW39630	Human polypeptide
22	367.5	6.1	749	22	AAW78509	Human protein SEQ
23	367.5	6.1	749	22	AAW78509	Human protein SEQ
24	367.5	6.1	788	22	AAU17163	Novel signal trans
25	367.5	6.1	777	22	AAW78508	Human protein SEQ
26	367.5	6.1	796	22	AAW79492	Human protein SEQ
27	367.5	6.1	796	22	AAW79493	Human protein SEQ
28	367.5	6.1	796	22	AAW41415	Human polypeptide
29	367.5	6.1	796	22	AAW41416	Human polypeptide
30	366	6.0	298	21	AAW20060	Arabidopsis thalia
31	364.5	6.0	244	21	AAW606211	Arabidopsis thalia
32	347.5	5.7	907	22	AAW78535	Human protein SEQ
33	344	5.7	250	21	AAW20061	Arabidopsis thalia
34	343.5	5.7	893	22	AAW79519	Human protein SEQ
35	336	5.5	557	22	ABW67606	Novel human diagno
36	334	5.5	272	21	AAW86335	Human gene 2-encod
37	328.5	5.4	519	22	AAW40360	Human polypeptide
38	328.5	5.4	533	22	AAW95321	Human protein sequ
39	326	5.4	371	22	AAW42146	Human polypeptide
40	325.5	5.4	457	22	ABW63679	Drosophila melanog
41	325.5	5.4	571	22	ABW67288	Drosophila melanog
42	325	5.4	1092	22	ABW20243	Novel human diagno
43	319.5	5.3	533	22	ABW60453	Human cell cycle a
44	315.5	5.2	363	22	ABW61447	Drosophila melanog
45	315	5.2	508	22	AAW93726	Human polypeptide

ALIGNMENTS

RESULT 1

AAV45096

ID AAV45096 standard; Protein; 1168 AA.

AC AAV45096;

XX 31-MAY-2000 (first entry)

DT Human TBC-1 protein.

DE TBC-1; human; blallelic marker; chromosome 4; cell cycle regulator;

KW tissue differentiation; yeast regulator; BUB2; cdc16; tre2-oncogene;

KW linkage analysis; genetic map; detection; diagnosis; genotyping;

KW mitosis; prostate cancer; transgenic animal; screening; cytokinesis.

XX Homo sapiens.

XX Key

XX Domain

FT 786...974

FT /label= TBC_domain

FT /note= "Regulates protein-protein interaction"

FT 886...893

FT /note= "This sequence interacts with a kinase"

XX WO200008209-A2.

XX 17-FEB-2000.

XX 06-AUG-1999;

XX 99WO-IB01444.

XX 07-AUG-1998;

XX 98US-0095653.

XX (GEST) GENSET.

XX

PI	Blumenfeld M, Bougueleret L, Chumakov I.
XX	WPI; 2000-205736/18.
DR	N-PSDB; AA250906, AA250907.
XX	New isolated human TBC-1 nucleic acids, useful for developing products
PT	for the diagnosis and treatment of disorders involving cell
PT	proliferation, particularly prostate cancer -
XX	Claim 16; Page 162-165; 166pp; English.
PS	The present amino acid sequence is the human TBC-1 protein, encoded by
CC	exons 1-2 and A-L. TBC-1 protein is involved in the regulation of cell-
CC	cycle, mitosis, cytokinesis and tissue differentiation in mammals. The
CC	TBC domain is homologous to regions in the tre2-oncogene and in the
CC	yeast regulators of mitosis BUB2 and cdc16. TBC-1 gene is mapped to a
CC	candidate region of prostate cancer on chromosome 4. An alteration of
CC	TBC-1 sequence may be associated with a pathological condition, resulting
CC	in abnormal cell proliferation leading to cancer, e.g. prostate cancer.
CC	Biallelic markers present in the sequence can be used for generation of
CC	genetic maps, linkage analysis and association studies. TBC-1 sequence
CC	can be used for detection, diagnosis, genotyping, production of
CC	transgenic animals and screening of compounds for use in therapy.
XX	Sequence 1168 AA:
SQ	Query Match 100.0%; Score 6055; DB 21; Length 1168;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 1168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MEPTFTARKHLLPNEVSVDFGLQLVGSPLVSHLTTPMLPVVAEVRRLSROSTRKEPV 60
DB	1 meptftarkhllpnevsvdfglqlvgsplvshlttpmlpvvaevrrlsrsgstrkepv 60
QY	61 TKQVRLCVSPSGLRCEPEGRSQOWDPLIYSIFECKPQRVKLIHNSHPDSYFACLIIK 120
DB	61 tkqvrlcvspsglrcepegrsqowdpliysifeckpqrvkhlnshdpdyfaclike 120
QY	121 DAVHRQSICVFRADDOOTYPEIISSTROAGTKARQEELHCPSFEFDTFESKKFEVLFCGR 180
DB	121 davhrsicvfraddootypeiisstroagtkarqeelhcpsfefdtfeskkfevlfcgr 180
QY	181 VTVAHKAPPALIDECIEKENHNHVSRSRPNPPHAAAPTQGEVPVRPMKRFSQPQ 240
DB	181 vtvahkappalideciekenhnhvrsrgsrpnpphaaptqgevpvrpmkrksfqpq 240
QY	241 LRSLAFKRELQDGLRRSGFPSSFEEDIBNLHLSGHNIVOPTDIENRTMLFTIGOSEV 300
DB	241 lrsلافkelqdglrrsgffssfeesdienhlisghnivgptdeenrtmiftigseve 300
QY	301 YLISPDTFKKTALENKFIESCSQIRHVPHFGFICRESSGGGGFHVCVVFQCTNEALV 360
DB	301 ylispdtkkialeknkfiescsqirhvphfgficressggggfhvcyvfqctnealy 360
QY	361 DEIMNTLKQAFTVAOVQATAKPAQLCEGPLOSLHKLCEIEGMNSKTLELQHLLTT 420
DB	361 deimntlkqaftvaovqtakapalcegcplgslhklceriegmnsktklelqhlltt 420
QY	421 LTNOEQATIFEVOKLRPRNEQRENELIISFLRCLYEEKQKEHIHGEMKOTSOMAANI 480
DB	421 ltneqeatifeevoklrprneqreneliisflrclyeekqkehiihgemkotsmaani 480
QY	481 GSELPSPATRFLDLNKNKAKRSITESTLESILSRGNKARGLOHSISVDLSLSLTSN 540
DB	481 gselppsatrfrldmlknknkaksiteslesilsrgnkarglohsisvdlsslstlsn 540
QY	541 TSKEPSVCEKALPISESSFKLKGSDDLSDSESHPPEEPAPLSPOOAFRRANTLSHF 600
DB	541 tskepsvcekealpisesfklkgssddlssdeshehppeepaplsppoafrrantlshf 600
QY	601 PIECQEPQPARGPSVGRKLMYHVSVTETPHERKDFESKANHLGDGSGTGVKTRRHS 660
DB	601 piecqepqpargpsvgvqrklmyhvsvtetpherkdfe skanhlgdsggtgvktrrhs 660

XX This is the amino acid sequence of a Tbc1 (tre-2, BUB2 and cdc16)
 CC polypeptide, isolated from a bone marrow-derived mast cell DNA library.
 CC The screen was carried out using a probe generated by a subtraction
 CC method which compared mRNA expression in an undifferentiated mast cell
 CC line P815 and cell line PC76 (P815 cells transfected to express the
 CC murine GATA-1 transcription factor - a factor which controls the
 CC expression of genes involved in mast cell differentiation). Tbc1 encodes
 CC a protein involved in the coupling of cell proliferation to cell
 CC differentiation, which can be used to treat leukaemia (especially acute
 CC myelogenous leukaemia) by causing leukaemic cells to differentiate.
 XX
 SQ Sequence 1141 AA;

Query Match 82.2%; Score 4976; DB 19; Length 1141;
 Best Local Similarity 87.9%; Pred. No. 0;
 Matches 970; Conservative 47; Mismatches 74; Indels 12; Gaps 5;

QY 37 MPMLPWWVAEVRRLSRQSTRKEPTVKQVRLCVSPSGRLRCEPEGRSQOWDPLIYSIFEC 96
 Db 1 mpmllpwvvaevrrlsggckkprtkqvrllwvpsgrlrcpdliekspwdplicsifec 60

QY 97 KPORVHLIHNHSDPSYFACILKEDAVHROSICVFKADDQTKVPEIISIRQAGKIARQ 156
 Db 61 kprvhlhlnhshdpsyfacklkdahrsqicvfkaddqtkvpelissirgagkiarq 120

QY 157 EELHCPSEFDFTSKFPEVLFCGRVTVVAHKKAPPALIDECIERFNHVSGRGSESPRNP 216
 Db 121 eelrcpsefdftfakkefvilfcgrvtvvhkappalidecierfhnvscgrtd----- 174

QY 217 PHAAPTGS-QEPVRRPMRKSFQPSGLSLAFRLKQGLRSGGFFSPESDIENHLIS 275
 Db 175 -weaptgqseaprpnrksfsgpgrlsrlafkfgdaslrss-tfsaf-dndienhlig 231

QY 276 GHNVQPTDEENRTMLFTTIGQSEVYLISPDTKKIALEKFKETSCQSGIRVHDHFGFI 335
 Db 232 ghnvqptdneenrtmlftlgpsevyllspdtkkialekfkaisfcsqgrhvdhfgfi 291

QY 336 CRESSGG--GGFHFVCVFOCTNEALVDEIMMTLKQFTVAAVQOATAKAPALCEGCLQ 393
 Db 292 crecsgggsgghfvcyvfqctnealvdeimmtlkqftvaavqgtakapalcegcplq 351

QY 394 SLHLKCEIRIEGMSNKKLEQLHLLTTLTNOEQATIFEVEQKLRPRNEORENELITSFLR 453
 Db 352 glhlcieriegmsnkkleqlhlltltltnoqatifeevqklrprneqrenelitsflr 411

QY 454 CLYEKQKEHIHIGEMKQTSQMAENIGSELPPSATFRDLMLKNAKRLSTESILS 513
 Db 412 clyeekqeshgtgaptqlqaenlgsdlppsasfrldsknkrakrlstesils 471

QY 514 RGNKARGLOEHSISVLDLSDSLSTLSNTSPKSPVCVEALPISESPKLLGSSDDLSSDS 573
 Db 472 rgnkarglqdasvdlssstlsntskelsmgkkaefpvsatsfkllgssddlsds 531

QY 574 ESHLPEEPAPLSPOAFRRRANTLSHPFIECQPPQARGSPGVSKLMRYHSVSTETP 633
 Db 532 eghlaeesallspqafrrrantlshfpvecpappespgvsgvrklmryhsvstetp 591

QY 634 HERKDFESKANHLGDSGTPVKVRRHSWROQIFLRVATPQACDSSSRDYEDYSELGLPP 693
 Db 592 herkdfeskanhlgttdgtvptrhswrqiflrvatpqaacdpsaryedyseelgealp 651

QY 694 RSPLEPCEDGPFPGPPPEEKRTSRRELRQWQKAILQOILLRMEKENQKLAASENDLLN 753
 Db 652 rsplepcedgpfpgykrgrthasfescgkrpsscslvrmekenqklqasendlln 711

QY 754 KRLKLYEETPCLEKVTWVKMLSTPGSKIKFDMKMHSAVQCQVPRHHRGEIWKFL 813
 Db 712 krlklyeetpclekvttwvkmstpgsrkikfdekvhsvavggvprhrhgeiwnkl 771

QY 814 AEQFLKHPSPSKOOPKDPVKELKOLTSQOHAILLDLGTRTPTPYFSAQLGAGLGL 873

Db 772 aeqfhkhkpfspkqpkdvpkykellkkltsqgqailldgrtftphpyfsaqigagqlsl 831

QY 874 YNLKAYSLLDQVGVCGQGLSFVAGIILLHMSBEEAFKMLKFLMDFMGLRQYRPDMIL 933
 Db 832 ynlkaysllldqvgvcqglsfvagillhmsseeafkmlkflmfdnglrkqyrdmll 891

QY 934 QIQMYQLSRLLDHYHRLDLYNHLHEEIGPSLYAAPWFLTMFASQPLGTFVARVDFMIFLQ 993
 Db 892 qiqmyqlsrlldhyrdlynhleehetgptyaapwfltfvasqplgtfvarvdfmiflq 951

QY 994 GTVIEFKVALSLGSHKPLILOHENLETIVDFIKSTLPNLGLVQMEKTIINQVPEMDIAKQ 1053
 Db 952 gsevfakvalsllgshkplilqhenletivdfikntlpnlglvqmektisqvfemdiakq 1011

QY 1054 LQAYEYEHVYQLBELIDSSPLSDNQRMKLEKTNSSLRKONLDELLEQLQVANGRIQSLA 1113
 Db 1012 lqayeveyhvqyeelliesplsdnqrmelektntslrkqldleqlqvanarigsla 1071

QY 1114 TIEKLSSSESKLQAMLTLELER 1136
 Db 1072 tveklitssesklqraitlever 1094

RESULT 3
 AAM38698
 ID AAM38698 standard; Protein; 937 AA.
 XX
 XX AAM38698;
 XX AC AC
 XX DT 22-OCT-2001 (first entry)
 XX DE Human polypeptide SEQ ID NO 1843.
 XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 XX PR 25-APR-2000; 2000US-0552317.
 XX PR 09-JUL-2000; 2000US-0598042.
 XX PR 19-JUL-2000; 2000US-0620312.
 XX PR 03-AUG-2000; 2000US-0653450.
 XX PR 14-SEP-2000; 2000US-0662191.
 XX PR 19-OCT-2000; 2000US-0693036.
 XX PR 29-NOV-2000; 2000US-0727344.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI57854.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 3; SEQ ID NO 1843; 10078pp; English.
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ

Query Match	79.7%	Score 4823	DB 22	Length 937
Best Local Similarity	100.0%	Pred. No. 0		
Matches 937	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	232	MRKFSQGLRLSIAFKELQDGLRSGFFSFESDIENHLISGHNIYQPTDIEENRTM	291	
Db	1	mrksfqgqlrsiafakelqdgglrsgffsfesdienhlisghniyqptdienrtm	60	
QY	292	LFTIGSEVYLLSPDKKTALENKFIEISFCSGIRHVDHFGFICRESSGGGFHVCVY	351	
Db	61	lftigsevyllspdkkaleknfkeisfcsgirhvdhfgficsressgggfhvfcyv	120	
QY	352	FOCTNEALVDEIMMTLKQFTVAQVQOTAKAPQACGEGPLQSLHLKCRIEGMSNKT	411	
Db	121	fqctnealvdeimmtlkqftvaavqatakapaqlcegcplqslhlcleriegmsnktk	180	
QY	412	LEIQKHLTLTNOEQATIEEVOKLRPRNEQRENEIIISFLRCLYBEKKEHIIHIGEMKO	471	
Db	181	leiqkhltiltnoqatifeevoklrprneqreneiiisflrclyeekkehiihigemkq	240	
QY	472	TSQMAAENIGSELPPSATFRDLMLKNAKRSITLESILSRGNKARGLOEHSISVDLD	531	
Db	241	tsqmaeenigselppsatfrldmlknkarsitesilsrgnkargloehsisvldld	300	
QY	532	SSLSSTLSNTSKBPSVCEKEALPISBSPKLLGSSDLSDSSESHLPPEPAPLSPQAFR	591	
Db	301	sslsstlntskbpsvcekealpisessfkllgssedlssdeseshlpeepaplsqqafr	360	
QY	592	RRANTLSHPPIECQEPQPARSGVQSKLMRYHVSVTETPHERKDFESKANHLGDSGG	651	
Db	361	rrantlshpniecqpqparsgvsgqrklmryhsvstetpherkdfeskanhlgdsgg	420	
QY	652	TPVKTRRHWSROOIFLRVATPOKACDSSRYEDYSELGELPPSPLEPCEDGPGPPPE	711	
Db	421	tpvktrrhswrqoiflrvatpqacdsrryedyseigelpsplepcedgpgpppe	480	
QY	712	EKKRTSRELWQKAILQOILLRMKENQKLOASENDLLNKKRLDYEEITPCLKEVT	771	
Db	481	ekkrtsrelwqkailqoillrmkenqkloasendllnkrldyeeitpclkevt	540	
QY	772	TWVEKMLSTPGRSKIFDMEKMSAVGQVPRHHRGEIWKFLAEQPHLKHQPPSKQPKD	831	
Db	541	twvekmlstpgrskikfdmekmsavgvprhrhrgewkflaegphlkhqfpskqpkd	600	
QY	832	VPYKELLKOLTSQOHAIDLGLTFTTPVFSQALGAGOLSLYNILKAYSLLDQEVGYCQ	891	
Db	601	vpykellkoltseqhaildglftftphvfsaqagqlslynilkayslldqevgycq	660	
QY	892	GLSFVAGILLHMSBEEAFKMLKFLMDFDMLGRKQYRPDMIIQIQYQISRLLDHYRDL	951	
Db	661	glsfvagillhmsbeefkmlkflmfdmldgrkqydpdmliiqyqisrllldhyrdl	720	
QY	952	YNHLEHEIGPSIYAPWFLTWPASOPPLGCFVARVDMFLQCTEIVFKVALSLGSHRP	1011	
Db	721	ynhleheigpsiyaapwflmtasdpigfvarvdmflqctevfkvalslgshkhp	780	

QY	1012	LILQHENLETIYDFIKSTIPNLGLVQMEKTIQVFPENDIAKQLQAYEVEYHVLQEEILDS	1071	
Db	781	lilqhenletiydfikstipnlglvqmektinqvfemdiakqlqayevyehvlqeelids	840	
QY	1072	SPLSNORMDKLEKTNSSRKONLDLEQLQVANGRIQSLEATIEKLLSESSEKQKQAMLT	1131	
Db	841	splsdnormdklektntsslrknldleqlqvangrigsleatiekllsesesklqamlt	900	
QY	1132	LELERSALQTVVEELRRRSAPSDREPECTQPEPTGD	1168	
Db	901	lelersallqtveelrrrsaeprsdrepectqpeptgd	937	
RESULT	4			
AAB43195				
ID	AAB43195	standard; Protein; 981 AA.		
XX				
AC	AAB43195;			
DT	08-FEB-2001	(first entry)		
XX				
DE	Human ORFX ORF2959	polypeptide sequence SEQ ID NO:5918.		
XX				
KW	Human; open reading frame; ORFX; detection; cytotstatic; hepatotropic;			
KW	vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;			
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;			
KW	hypnotic; thrombolytic; coagulant; vasotropic; antidiabetic;			
KW	immunostimulant; dermatological; immunosuppressive; antiinflammatory;			
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;			
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;			
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;			
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;			
KW	cholesterol ester storage; systemic lupus erythematosus; infection;			
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;			
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;			
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;			
KW	thrombosis; contraceptive.			
XX	Homo sapiens.			
OS				
XX	WO2000058473-A2.			
PN				
XX	05-OCT-2000.			
PD				
XX	31-MAR-2000; 2000WO-US08621.			
PF				
XX	31-MAR-1999; 99US-0127607.			
PR	02-APR-1999; 99US-0127636.			
PR	05-APR-1999; 99US-0127728.			
PR	30-MAR-2000; 2000US-0540763.			
XX	(CURA-) CURAGEN CORP.			
PA				
XX	Shinkets RA, Leach M;			
PI				
XX	WPI; 2000-602362/57.			
DR	N-PSDB; AAC77404.			
DR				
XX	Novel nucleic acids and peptides derived from open reading frame x,			
PT	useful for treating e.g. cancers, proliferative disorders,			
PT	neurodegenerative disorders and cardiovascular disease -			
XX				
PS	Claim 11; Page 5097-5099; 5507pp; English.			
XX				
CC	AAC74446 to AAC77606 encode the proteins given in AAC40237 to AAC43397,			
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX			
CC	sequences have activities such as: cytotstatic; hepatotropic; vulnary;			
CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;			
CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;			
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;			
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;			
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;			

antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 981 AA;

Query Match 78.8%; Score 4772; DB 21; Length 981;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 929; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

Qy 178 CGRTVAHKAPALIDIEIKFNHVSGRSGSESPRNPHPAFTGSEPPVRRMRKSF5 237
Db 2 cgrvtvahkappalideciekfhnvsgsgsesprnpnpphaarhregpvrrpmrkfs 61
Qy 238 QPGLRSAPFRELODGLRSGFPSSPEESDIENHLISGHNIVOPTDIEENRTMLFTIGQ 297
Db 62 qpgrlsafrelqdggrsgrfssfeesdienhlisghnivaqtdieenrtmlftigq 121
Qy 298 SEVYLISPDTKKIALEKNFKIESCSQGIHVHDFGFTICRESSGGGFHFVYVFOQTNE 357
Db 122 sevylispsdtkkialeknfkiescsggihrvhdfgfticressggghfvvyfqtne 181
Qy 358 ALVDEIMMTLQAFVAAVOQTAKAPALCEGCPQLSLKLCERIEGNMSSKTKLELQKH 417
Db 182 alvdeimmtlkqafvvaavqfakapalcgcpqlslhklceriegnmssktklelqkh 241
Qy 418 LTTLTNQOANIFEVQKLRPNORENELIISFLRCLYEKQKEHIIHIGEMKOTSOMAA 477
Db 242 lttltngqatifeevqklrpnoreneliisflrclyeqkehehiihigemkotsmaa 301
Qy 478 ENIGSELPSPATRRFLDMKKNKRSLTESLISLRGNKARGLOEHSISVDLSSLSST 537
Db 302 enigselppsatrrfldmknkarslteslislrgnkargloehsisvdlsslsst 361
Qy 538 LSNTSKEPSVCEKALPISSESKLLGSSEDLSSDSESHLPEEPAPLSPQAFRRANTL 597
Db 362 lsntskepsvcekalpisessefklgssedlssdseeshlpeepaplsppqafrrantl 421
Qy 598 SHFPIEQEPPQARGSPGVQSKLMRYHSVSTETPHERKDFESKANHLGDSGCTPVKTR 657
Db 422 shfpieqepqpargspgvqsklmryhsvstetpberkdfeskanhlgdsgctpvktr 481
Qy 658 RHSNRQITFLAVTPQKACDSSRYEDYSELGELPPSPLEPCVEDGPGFPPEEKRTS 717
Db 482 rhsnrqitflavtpqkacdssryedyselegelppsplepcedgpgfppeekrts 541
Qy 718 RELRELQKATLQOILLIRMEKENQKLAQSENLLNKLRLKLDYEITPCIKREVTVWEMK 777
Db 542 vssescgkiffrnrycxlrmekenqklaqsenllnklrlkldyeitpcikrevttvekm 601
Qy 778 LSTPGRSKIKFDMKMSAVGGVPRHHRGEIWKFLAEQFHLKHQFPFSKOQKDPVYKEL 837
Db 602 lstpgrskikfdmkmsavggvprhrhrgelwkflaeqfhlkhqfpskqgkdpvykel 661
Qy 838 LKQITSOQHALIDLGTFTFHPYFSQALGNQSLNKLKAYSLLDQEVGYCGGLSFVA 897
Db 662 lkqitsoqhalidlgtrftfchpyfsaqalgnqslnklkayslldqevgycgglsfva 721
Qy 898 GILLHMSSEAFKMLFLMDFMGLRQKRPDMIILOITQMYQLSRLLHRYDRDLYNHL 957
Db 722 gillhmseefkmlflmfdmglrkqrpdmililqmyqlsrllhrydrdlynhlee 781

Qy 958 HEIGPSIYAAPFWITMTFASQPPFLGVFARVDFMIFLOQTEVIFKVALSLGLSHKPLIQHE 1017
Db 782 heigpsiyaapfwitmtfasqppflgvfardfmdiflqgtevifkvalsilgshkpliqhe 841
Qy 1018 NLETIVDFIKSTLNLGLVOMEKTINGVFMEDIAKQIQAYEYHVHQEELIDSSPLSDN 1077
Db 842 nletivdfikstlnlglvomektingvfmediakqiqayeyhvhaqeelidssplsdn 901
Qy 1078 QRMDEKLEKTSRLKQNLDDLEQLQVANGRIQSLEATIEKLSSSKLKQAMLTLELERS 1137
Db 902 qrmdekletssrlkqnldlleqlqvangriqsleatieklssesklqamltlelers 961
RESULT 5
AAM40484
ID AAM40484 standard; Protein; 763 AA.
XX
AC AAM40484;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5415.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
WI: 2001-442253/47.
N-PSDB; AAI59640.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX novel as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5415; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI51369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 763 AA;

Query Match	64.3%	Score 3891;	DB 22;	Length 763;
Best Local Similarity	99.7%	Pred. No. 0;		
Matches 758;	Conservative	2;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy 409	KTLELQKHLTTNQDQATFEVQKLRPRNEORENELIISFLRCLYEKQKEHIGI	468		
Db 4	ktlelqkhlittnqeqatfeevqklrprneoreneliisflrcifeekqkehigie	63		
Qy 469	MKQTSQMAENIGSELPSPATRFRLMLKNKAKRSLTESLESILSRGNKARGLOEHSISV	528		
Db 64	mkqtsqmaenigselppsatrfldmlknkaksrleslesilsrngnkarglqehaisv	123		
Qy 529	DLDSLSSTLSNTSKPSVCEKEALPISSESKLLGSSEDLSSDSHLPPEEPALSPQO	588		
Db 124	dlsslsstlsntskpsvcekealpisesskllgsessedlssdseshlpeepalpspq	183		
Qy 589	AFRRRANTLSHPFIECOEPPQARGSPGVQSQRKLMRYHSVSTETPHERKDFESKANHLGD	648		
Db 184	afrrrantlshflecpeppqargspgvqqrklmryhsvstetpherkdfeskanhlgd	243		
Qy 649	SGGTPVTRRHSWQOJFLRVATPQKACDSSSRDYSELGELPPRPLPVCEDGFGFP	708		
Db 244	sggtpvtrrshwqjflrvatpqkacdssrydyselgelpprplpvcvedgfgfp	303		
Qy 709	PPEKKTSTRELRLQKAILQIILLRMKKNOKLOASENDLNLKRLDYETITPCLK	768		
Db 304	ppekktstrelrlwqailqillrmekengkldasendlnlrlkldyeeitpclk	363		
Qy 769	EVTTVWEKMLSTPGRSKIKEDMEKMHSAVGQVPRHHRGEIWKFLABQFHLKHQFSPKQO	828		
Db 364	evttvwekmlstpgsrskikfemkmsavggvprhhrgeiwlkflaegfhlkhqfspkq	423		
Qy 829	PKQVPYKELLKQTSQOHAIIIDLGRFPFPPYPSAOLGAGOLSLYLKAYSLLDQEVG	888		
Db 424	pkdvpkellkqltsqghailidgrtftphysaqlgagqlslynlkayslldqevg	483		
Qy 889	YCQGLSVAGILLRMSSEAFKMLFMFDMGLRKQYRPMIILQIQMLSLLDHYH	948		
Db 484	ycqglsvagillrmseeeafkmlfmdgmglrkqyrdpmiilqmqvqlslrldhyh	543		
Qy 949	RDLYNHLEEHIGPSLYAAPWFLTMFASQPLGPFVARVDMIFLQGTVEVIFKVALSLGS	1008		
Db 544	rdlynhleehigpslyaaawfltmfasqplgpfvarvdmiflqgtvevifkvalsllgs	603		
Qy 1009	HKPLILOHENLETIVDFIKSTPLNGLVQMEKTIQVFMENDIAKQLOAYEYHVHLOEEL	1068		
Db 604	hkpilohenletivdfikstplnqlvqmektiqvfmendiaqlqayevyhnvloeel	663		
Qy 1069	IDSSPLSDNQRMDKLENTSNLRKQNLDLLEQLQVANGRIQSLEATIEKLSSSKLKQA	1128		
Db 664	idsplsdnqrmdklekntsnlrkqnlldleqlqvangriqsleatieklssesklkqa	723		
Qy 1129	MLTLELERSALLQVFEELRRRSAPSDREPECTQPTGTD 1168			
Db 724	mtlelersallqtveelrrrsakpsdrepectqptgd 763			

RESULT 6
 AAB95675
 ID AAB95675 standard; Protein; 674 AA.
 XX
 AC AAB95675;
 XX
 DT 26-JUN-2001 (first entry)
 XX

DE Human protein sequence SEQ ID NO:18466.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 OS
 PN EP1074617-A2.
 XX
 XX 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 18466; 2537pp + CD ROM; English.
 PS
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 674 AA;

Query Match	57.2%	Score 3463;	DB 22;	Length 674;
Best Local Similarity	100.0%	Pred. No. 5.7e-269;		
Matches 674;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy 495	MLKNKAKRSLTESLESILSRGNKARGLOEHSISVYDLSSLSSTLSNTSKPSVCEKEALP	554		
Db 1	mlknkaksrleslesilsrngnkarglqehsisvldsslsstlsntskpsvcekealp	60		
Qy 555	ISESSFKLLGSSEDLSSDSHLPPEEPALSPQOAFRRRANTLSHFFIECOEPPQARGS	614		
Db 61	isessfkllgsessedlssdseshlpeepalpspqafrrrantlshffiecoepqargs	120		
Qy 615	PGVYSQRKLMRYHSVSTETPHERKDFESKANHLGDSGGTTPVKTRRHSRQOJFLRVATPQK	674		
Db 121	pgvysqrklmryhavstetpherkdfeskanhlgsdggtpvktrrhswrqjflrvatpqq	180		

Qy	675	ACDSSRYEDYSELGELPPRSPLEPCVCDGPGPPPEKKRTSRELRLWQAILQOILL	734	PR	07-JUL-2000;	2000US-0216647.
Db	181	acdsrsryedyseigeippsrplepvcvedgpgpppekkrtrelrelwqallqill	240	PR	07-JUL-2000;	2000US-0216880.
Qy	735	LRMEKENQKLOASENDLLNKKLKDYEIITPCLEKVTIVWEKMLSTGPRSKIKFDMKMH	794	PR	11-JUL-2000;	2000US-0217487.
Db	241	lrmekenqkloasendllnrklkidyeeitpclekvttvwekmlstgprskikfmekeh	300	PR	11-JUL-2000;	2000US-0217496.
Qy	795	SAVGQVPRHHRGIIWIFLAQFHLKHQFPKQOPKDPYKELKQLTQQHAILIDLGR	854	PR	14-JUL-2000;	2000US-0218290.
Db	301	saveggvprhrgeiwiwiflaeqfhlkhqfksqpkdvpkyellkqltsqqhailidgr	360	PR	26-JUL-2000;	2000US-0220963.
Qy	855	TFPTHYFSAQAGQSLYNILKAYSLLDQEVGYCGSLFVAGILLHMESEEAFAKMLK	914	PR	26-JUL-2000;	2000US-0220964.
Db	361	tftphpyfseqagqislynilkayslldqevgycgslfvagillhmseeeafkmlk	420	PR	14-AUG-2000;	2000US-0224518.
Qy	915	FLMFDGMLRKQYRPMILQITQWQLSRLLHDYHRLDLYNHLBEHIGPSLYAAPFLTMF	974	PR	14-AUG-2000;	2000US-0224519.
Db	421	flmfdmglrkqyrdpmilqimvqlsrllhdyhryldynhlbeheigpslyaaapfltmf	480	PR	14-AUG-2000;	2000US-0225213.
Qy	975	ASQPLGFVARVDFMIFLQGTVEFIKVALSLGSHKPLILQHENLETIVDFIKSTLPNLG	1034	PR	14-AUG-2000;	2000US-0225214.
Db	481	asqplgfvavrfmiflqgtvefivkvalslgshkplilqhnenletivdfikstlpnlg	540	PR	14-AUG-2000;	2000US-0225266.
Qy	1035	LVOMEKTINQVFEMDIKQLOAYEVEYHVLQELIDSSPLSDNQRMDKLEKTNSSLRKQN	1094	PR	14-AUG-2000;	2000US-0225757.
Db	541	lvomektinqvfemdiekqlqayevyhlqeelidssplsdnqrmdklektnsslrkqn	600	PR	14-AUG-2000;	2000US-0225758.
Qy	1095	LDLLEQLOVANGRIOSLEATEIKLSSSKLKQAMLELERSALLQTVBELRRRSAPPS	1154	PR	18-AUG-2000;	2000US-0225759.
Db	601	ldlleqlovangriqslsleatiekllssesklkqamltlelrsallgtveelrrrsapps	660	PR	18-AUG-2000;	2000US-0226279.
Qy	1155	DREPECTQPEPTGD 1168		PR	22-AUG-2000;	2000US-0226681.
Db	661	drepectqpeptgd 674		PR	22-AUG-2000;	2000US-0226681.
RESULT 7						
ID	AAU21588					
XX	AAU21588 standard; Protein: 343 AA.					
AC	AAU21588;					
XX	04-DEC-2001 (first entry)					
XX	Novel human neoplastic disease associated polypeptide #21.					
DE	Human; neoplastic disease associated polypeptide; cancer;					
KW	hyperproliferative disorder; neural disorder; immune system disorder;					
KW	muscular disorder; reproductive disorder; gastrointestinal disorder;					
KW	pulmonary disorder; cardiovascular disorder; renal disorder;					
KW	neuroprotective; cytostatic; anti inflammatory; vasotropic.					
XX	Homo sapiens.					
OS	XX					
XX	WO20015163-A1.					
XX	02-AUG-2001.					
XX	17-JAN-2001; 2001WO-US01358.					
XX	31-JAN-2000; 2000US-0179065.					
XX	04-FEB-2000; 2000US-0180628.					
XX	24-FEB-2000; 2000US-0184664.					
XX	02-MAR-2000; 2000US-0186350.					
XX	16-MAR-2000; 2000US-0189874.					
XX	17-MAR-2000; 2000US-0190076.					
XX	18-APR-2000; 2000US-0198123.					
XX	19-MAY-2000; 2000US-0205515.					
XX	07-JUN-2000; 2000US-0209467.					
XX	28-JUN-2000; 2000US-0214886.					
XX	30-JUN-2000; 2000US-0215135.					

20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 08-NOV-2000; 2000US-0249207.
 08-NOV-2000; 2000US-0249208.
 08-NOV-2000; 2000US-0249209.
 08-NOV-2000; 2000US-0249210.
 08-NOV-2000; 2000US-0249211.
 08-NOV-2000; 2000US-0249212.
 08-NOV-2000; 2000US-0249213.
 08-NOV-2000; 2000US-0249214.
 08-NOV-2000; 2000US-0249215.
 08-NOV-2000; 2000US-0249216.
 08-NOV-2000; 2000US-0249217.
 08-NOV-2000; 2000US-0249218.
 08-NOV-2000; 2000US-0249245.
 08-NOV-2000; 2000US-0249245.
 08-NOV-2000; 2000US-0249264.
 08-NOV-2000; 2000US-0249265.
 08-NOV-2000; 2000US-0249297.
 08-NOV-2000; 2000US-0249299.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI: 2001-465558/50.
 N-PSDB; AAS34787.
 Novel polypeptides and polynucleotides useful as diagnostic reagents to
 diagnose diseases or disorders associated with aberrant expression or
 activity of polypeptides, and for treating cancers, rheumatoid
 arthritis
 Claim 11; SEQ ID No 315; 687pp; English.
 The present invention relates to the isolation of novel human neoplastic
 disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
 sequences encoding for these polypeptides. The sequences of the
 invention are useful in the diagnosis, treatment, prevention and/or
 prognosis of disorders involving neoplastic disease such as
 hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder

CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
 CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
 CC also be useful for treating other disorders such as neural disorders,
 CC immune system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC and renal disorders. The polynucleotide sequences of the invention are
 CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
 CC neoplastic disease associated polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 343 AA;
 Query Match 26.4%; Score 1600.5; DB 22; Length 343;
 Best Local Similarity 96.4%; Pred. No. 7.1e-120;
 Matches 320; Conservative 2; Mismatches 5; Indels 5; Gaps 1;
 QY 837 LKQLTSQOHAILDILGRTPPTHPYSAOLGAGQLSLYNILKAYSLLDQEVGYCQGLSFV 896
 DB 17 lllql-----icilftgrtftphysaqlgagqlslynilkayslldqevgycqlsfv 71
 QY 897 AGILLHMSEEEAFKMLKFLMFDGMLRKQYRPMIILQIOMYOLSLRLLDHYHRLDLYNHLE 956
 DB 72 agillhmseeeafkmlkflmfdmglrkqyrdmllqmqylslrllhdyhrdlynhle 131
 QY 957 EHEIGPSLYAAPFLTMEASORFPLGFVARVDFMIFLQGTVEIKFVALSLGSHKPLILQH 1016
 DB 132 eheigpslyaaapfltmfasqfplgfvarvdfmiflqgtveikfvalsllgshkplilqh 191
 QY 1017 ENLETIVDFIKSTPLNLGLVQMEKTIQVFMEDIAKQAYEYHVQLBELIDSSPLSD 1076
 DB 192 enletivdfikstplnlglvqmektiqvfemdiaqkqayevyhlqelidssplsd 251
 QY 1077 NORMDKLEKTNSSRLKONLDLLEQLQVANGRIOSLEATEKLLSSSKLKQAMLTLELER 1136
 DB 252 ngrmdklektnssrlkqnlldleqlqvangrigsleatiekllssesklqamltleler 311
 QY 1137 SALLQTVELRRBSAPSDREPECTOPEPTGD 1168
 DB 312 sallqtveelrrbsapdsrepectptgd 343
 RESULT 8
 AAU21586
 ID AAU21586 standard; Protein; 390 AA.
 XX
 AC AAU21586;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Novel human neoplastic disease associated polypeptide #19.
 XX
 KW Human; neoplastic disease associated polypeptide; cancer;
 KW hyperproliferative disorder; neural disorder; immune system disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW neuroprotective; cytostatic; anti inflammatory; vasotropic.
 XX
 OS Homo sapiens.
 XX
 PN WO200155163-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01358.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 22-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-465558/50.

XX N-PSDB; AAS34785.

XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, and for treating cancers, rheumatoid
 PT arthritis

PS Claim 11; SEQ ID No 313; 687pp; English.


```
OY 558 SSFKLL--GSSEDLSSSESHLPREPAPLSPOQAFRRRANTLSHPTECQEPPOPARGSP 615
Db 554 hnlrdiregsaeplgtq-----spgefrrsntvgasp-----sskp 591
OY 616 GVSQRK--LMRYHSVSTPHERKDFESKANHLGDSGTPVKTTRHSRQOIFLRVATPQ 673
Db 592 taqlkspmdiflkvgnspk-----aethgswrqailnsvtpps 633
OY 674 KACSSSRDYEDYSELGLPPRSPLEPVCEDPGPPPEKRTSRRELWOKAILQOIL 733
Db 634 kgldegyptefls-----pmckpakrgkrdaaelwrtairqtim 675
OY 734 LLRWEKENQKLOASENDLLNKRKLKLDVEETPCLEKVTWVEKML---STP-GRSKTKFD 789
Db 676 lnrmetenamlqarqnenelkrldklyeeivpcdkqlerweqiernstqignk---d 732
OY 790 MEKMSAVGQGVPRHHRGEIKWFLAEQPHLKHQFP--SKQPK-DVPYKELLKOLTSQOH 846
Db 733 pkvighairtyprskgvdvtfiaeq-hsmntapvdkrfpnfntpyhmlklhtebqh 791
OY 847 AILDGLGRTFTTHFYFSAQLGAGOLSLYNILKAYSLLDQEVGYCOGLSFVAGILLHMS 906
Db 792 aifidlgtrfphgkydplglglqlslfnllkayslldpelgycgglgfcgvlllhcd 851
OY 907 EEAFFKMLKFLMDGLKQYRPDMIILOIQMYQLSRLLHDYRDLNHLHEEIGPSLYA 966
Db 852 ansqglkhlmfrnmrkyldmknkqglqylqslrvkxhlpdlyvldqndvsptiya 911
OY 967 APWELTFASQFPGLGFVARVDFMIFLQGTVEIFKVALSLLGSHKPLILQHENLETIV 1026
Db 912 apwiltvfssqfpgfvarvdfllflessvifkfaiallsavhqqllakndfeelm 971
OY 1027 KSTLPNGLVOMEXTINQVFMENDIAKQLOAYEVEYHVLQEBLIDSSPLSDNORMDKL--E 1084
Db 972 ktvypkmehtcmeglmklvfmdgkqlaeqkveynvlqesi-----ttnhhlemnre 1026
OY 1085 KTNSSLRKQNDLLEQLQVANGRIQSLEATIEKLLSESRLKQAMLTLELERSALLQTV 1144
Db 1027 kt-----qnhleqqqlfagslaqlett-----rssqqaqltlqsgvqseletiq 1073
OY 1145 ELRRRSAPSDREPECTOP 1163
Db 1074 tlgyvgqlvehnpdlelp 1092

RESULT 10
AAM93840
ID AAM93840 standard; Protein; 265 AA.
XX AC AAM93840;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide, SEQ ID NO: 3915.
XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX OS Homo sapiens.
XX PN EPI130094-A2.
XX PD 05-SEP-2001.
XX PF 07-JUL-2000; 2000EP-0114089.
XX PR 08-JUL-1999; 99JP-0194486.
XX PR 11-JAN-2000; 2000JP-0118774.
XX PR 02-MAY-2000; 2000JP-0183765.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
```

```
XX WPI; 2001-524255/58.
DR N-PSDE; AAK94796.
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX Claim 8; SEQ ID NO 3915; 1380pp + sequence listing; English.
XX The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX SQ Sequence 265 AA;
```

```
Query Match 22.0%; Score 1334; DB 22; Length 265;
Best Local Similarity 100.0%; Pred. NO. 1.2e-98;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 904 MSEEAFKMLKFLMDGLKQYRPDMIILOIQMYQLSRLLHDYRDLNHLHEEIGPS 963
Db 1 mseeafkmlkflmdglrkqyrdmliiqmyqlsrllhdyrdlynhleeheigps 60
OY 964 LYAAPWELTFASQFPGLGFVARVDFMIFLQGTVEIFKVALSLLGSHKPLILQHENLETIV 1023
Db 61 lyaaapwiltvfssqfpgfvarvdfmiflqgtveifkvalsllgshkplliqhenletiv 120
OY 1024 DFIRSTLPNGLVOMEXTINQVFMENDIAKQLOAYEVEYHVLQEBLIDSSPLSDNORMDKL 1083
Db 121 dfirstlpnlglvomextinqvfemdiakqlayevyehvlqeeldssplsdnqrmkl 180
OY 1084 EKTNSLRKQNDLLEQLQVANGRIQSLEATIEKLLSESRLKQAMLTLELERSALLQTV 1143
Db 181 ektnsslrkqndlleqlqvangriqsleatiekllssesrlkqamltlelersallqtv 240
OY 1144 EELRRRSAPSDREPECTOP 1168
Db 241 eelrrrsapdsrepectqptgd 265
```

```
RESULT 11
AAM40436
ID AAM40436 standard; Protein; 1118 AA.
```

```
XX AC AAM40436;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 5367.
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia.
XX OS Homo sapiens.
XX PN WC200153312-A1.
XX PD 26-JUL-2001.
```

PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 23-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AAI59592.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2; SEQ ID NO 5367; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAI38642-AAI42213) with neurotropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities, chemotactic/chemokinetic activity, haemostatic
 CC activin/inhibin activity, cancer diagnosis and therapy, drug screening,
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1118 AA;

 Query Match 9.7%; Score 588.5; DB 22; Length 1118;
 Best Local Similarity 24.5%; Pred. No. 8.2e-38;
 Matches 229; Conservative 137; Mismatches 311; Indels 257; Gaps 32;

 QY 447 LIISFLRCLYEKQKCHIGEM---KOTSMAAENIGSELPPSATFRFLDM-----L 496
 DB 68 Llvkllwfpfgkcmrhhlqimevmrkqdsrivngnseqqlqkeladvimdpmdqpgge 127
 QY 497 KNAKRSLETES-----LESILSGNKAR-----GLQEHISIVDL-----DSSLSTPLSNT 541
 DB 128 kelvkrsqldgedgplnqlsastinpvlvlgldkpmisipkvpggdsasppftpv 187
 QY 542 SKPSV-----CEKEAL-----PISESSFKLL 563
 DB 188 adedsvvfkltlylgsavnaprsevealrmmsilrsqcqslsvtvlsvpnvsegivrl 247
 QY 564 -----GSSE-DLSSDSESHLPEEPAPL----- 584
 DB 248 dpqntelanypylkfcvrgdtpesdcftateshynaelfrihvfrcelqavsr 307
 QY 585 --SPOQAFRRANTLSHFIEQEPPOPGARGSGVQSKMLRYHVSSTPHERKDFESK 642
 DB 308 lysafafirsa-----kqtplsataaptpdsdlftf-svslel-----kddgk 352
 QY 643 ANHLGSGGTPVKTRHSMR-----QQIFLRVATPK----- 674
 DB 353 ----gyfsavpkdkdrqcfkrlrgldkklivvqqttnkelaiercfcglllspgkdvrns 408

QY 675 -----ACDSSRYEDYSELGELPRSP-LEPVCGDGPFGP----- 708
 DB 409 dmhlldlesmgkssdgsyvitgswnpkspghfvvneetpkdkvlftmtavdlvitevqe 468
 QY 709 -----PPBEK-----KRTSRE-----LRELWQKA-----ILOQILLRME 738
 DB 469 pvrflletkvrvcspnerlfpwfskrsttenffklkqikqrerkntdtlyevvccese 528
 QY 739 KENQKLO-----ASENDLLNKLKLDYEE-----ITPCLKEVTV 773
 DB 529 sererrkttaspsvrlpgsgsgssvipsdpdedeendepllsgsgdsvskecaekilet 588
 QY 774 WEKMLSTPGSRKTKFDMEKMHSAVGOGVPRHHRGETWTKFLAEQFHLKHOPPSKQPKDVP 833
 DB 589 wgeils-kwhlnlnvprkqlsslvrgvpealrgevwwqllagchndhilev----- 639
 QY 834 YKELLKQLTSQQHAILDGLRTPPTPHYPSAQLGAGQLSLYNILKAYSLLDQVGVYCOGL 893
 DB 640 yrillckespqdsaitrdinrtfphadyfkdtdgddgslykicakysvyydeesigycqgq 699
 QY 894 SFVAGILLHMSSEAEAFKMLKFLMDFMGLRKQYRPMIILQIQMYQLSRLLLHHDLYN 953
 DB 700 sflaavlllhmpdeeqafsvlvmfdyglreifkqnfedlhckfyqlerlmgeyipdlyn 759
 QY 954 HLEEHEIGPSLYAAPWELTMEFASORPLGFVARVDFMIFLQGTVEIFKVALSLGSHKPLI 1013
 DB 760 hfldislaehmyasqwlifltakfplymfvhlidlllceglsvfnvalglktkdddl 819
 QY 1014 LOHENLETIVDFTKSTPLNLGLVQMEKTTINQVFE-----MDIA-KIQOAYEVEVHYQEL 1068
 DB 820 ll-tdfegalkfrfvqlpk--tyrseenakkimelacnmkiskqkkkkyekyhtmtreq 876
 QY 1069 I-DSSPLSDNQMD-KLEKTNSSLRKQNLLEQLQVANGRTOSLEATIEKLSSSEKLUK 1126
 DB 877 aqgedpiferenrrrlqeanrlegenddlahel-----vtskialrkdldnaeakd 930
 QY 1127 QAMLTLELRSALLQVTEELRRRSAPSDREPEC 1160
 DB 931 alnkellmtkqklidaeeekrrleesahlkkmk 964

 RESULT 12
 AAM38650
 ID AAM38650 standard; Protein; 1069 AA.
 XX
 AC AAM38650;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 1795.
 KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukaemia.
 XX
 OS Homo sapiens.
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 PF 26-DEC-2000; 2000WO-US34263.
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;
WPI: 2001-442253/47.
N-PSDB; AAI57806.
Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
Example 3; SEQ ID NO 1795; 10078pp; English.
The invention relates to human nucleic acids (AAI57798-AAI61369) and
the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders.
Note: The sequence data for this patent did not form part of the printed
specification.
Sequence 1069 AA;

Query Match 9.68; Score 580; DB 22; Length 1069;
Best Local Similarity 23.98; Pred. No. 3.7e-37;
Matches 221; Conservative 137; Mismatches 311; Indels 254; Gaps 33;
295 IGQSEVYLISPTDKIALEKFKELSCGQIRHVDHFG-----FICRESSGGGPHFV 348
191 vsegvrlldpqtetanyplkfcvrg-----hdgtpesdcfafteshynaefr 244
349 CYVQCTNEALVDEIMTLKQAFVAAVQOPAKAPQCEGCPQLSHKLCIEREGMNS 408
245 ihvfceqlqavsrllsfafarsaktpisata--apqtpsdiftfsvsle----- 297
409 KTKLELQKHLTLNQEQTIFEEVQKLRPRNEORENELIISFLRC--LYEEKQKEHIH 466
298 -----ikeddkgvysav-----pkddrg-----cfklrggldkklviy 332
467 GEMKQTSOMAENIGSELPPSATFRFLDMLKNKAKRSLSLESILSRGNKARGLQEHSI 526
333 vqqttnkeialerc-----fgl-----llspgkdvrsdmhll 365
527 SVDLSSLSLTSNTSKPEVCEKALPISBSFKLLGSSDLSSSHL-----PEE 580
366 --dies-----mgk-----ssdgksyvitgswpks 389
581 PAPLSPQAFRRANTLSSHPIEQEPPQPPARGSPGVQKLMRYHSVSTETPHERKDFE 640
390 P-----hfqv-----vneetpkdkvlfm 407
641 SKANHLGSGGTPVKTRRHSRQOIFLRVATPQKADSSSRVEDYSELGELPPRSPLEPV 700
408 ttavdl-----vlt-----evqevrfiletkvrvcsperl----- 439
701 CEDGPFPGPPPEKRTSRE-----LRELWQKA-----ILQOILLRMEKENQKLQ 745
440 -----fwp-----fskrsttenffklkqkrknttdlyevvclesesererkrtas 491
746 -----ASENDLLNKLKLDYEE-----ITPCLKEVTVWEKMLSTPGRS 784

Db 492 psvrlpgsgsqsvipppdeedndedplllsgsgdsvskeacaeakiletwgeilis-kwhl 550
Qy 785 KIKFDMKMHSAVQGVPRHRHGEIWKFLAQPHLKHQFPKQKQPKDVPYKELLKQLTSQ 844
Db 551 nlrvpqlselsvngypealrgvqllagchndhivke-----yrillitkespq 602
Qy 845 QHAILIDLGRTPPHYFSQALGAGQSLYNILKAYSLLDOEVGYCQGLSFVAGILLIHM 904
Db 603 dsaltrdnrtfpahdyfkdgtggdgslykckaysvydeeligcvgqgsflaavillhm 662
Qy 905 SEEEAFKMLFLMDGLMGRKQYRPMIILQIQMQLSRLLDHYHRLDLYNHLHEEIGPSL 964
Db 663 peeqafsvlvkfmfdyglrelfkqnfedhckfyqlerlmqeyipdlynhfdlsleahm 722
Qy 965 YAAFWFTMFASQPFPLGFVARVDMIFLQGTVEVIFKVALSLGSHKPLILQHENLETIVD 1024
Db 723 yasqwlftlftakfplymvfhldllcegisvifnvalgllktskddlll-tdfegalk 781
Qy 1025 FIKSTLNLGLVQVEKINQVFE-----MDIA-KOLOAYEVHYVHLOEELI-DSSPLSDNQ 1078
Db 782 ffrvqlpk--ryrseenakkimelacnmklsqkklkkyekyhtmreqgaqgedpferfe 839
Qy 1079 RMD-KLEKTNSSLRKQNLDLLEQLQVANGRIQSLEATIEKLLSSESKLKQAMLTELEPRS 1137
Db 840 renrlqeammrlqegenddlahel-----vtskialrkdldnaeeekadainkellmtkq 893
Qy 1138 ALLQTVBELRRRSAPPSDREPEC 1160
Db 894 klidaeeekrrleesaqlkmc 916
RESULT 13
AAB68892
ID AAB68892 standard; Protein; 1069 AA.
XX AC AAB68892;
XX DT 24-APR-2001 (first entry)
XX DE Human RECAP polypeptide, SEQ ID NO: 22.
XX KW Human; RECAP; receptors and associated proteins; cerebroprotective;
XX KW neurotropic; neuroprotective; anticonvulsant; antiparkinsonian; anti-HIV;
XX KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;
XX KW antithyroid; immunosuppressive; nephrotropic; antigout; thyromimetic;
XX KW cytostatic; antibacterial; virucide; fungicide; protozoacide;
XX KW antiarteriosclerotic; hepatotropic; gene therapy; infection; cancer.
XX OS Homo sapiens.
XX PN WO200107612-A2.
XX PD 01-FEB-2001.
XX PF 21-JUL-2000; 2000WO-US20035.
XX PR 21-JUL-1999; 99US-0145232.
XX PR 07-OCT-1999; 99US-0158578.
XX PR 12-NOV-1999; 99US-0165192.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Au-Young J, Bandman O, Tang YT, Yue H, Azimzal Y, Burford N;
XX PI Baughn MR, Lu DAM, Hillman JL, Patterson C, Lal P;
XX DR WPI: 2001-168554/17.
XX DR N-PSDB; AAF58616.
XX PT Novel receptors and associated proteins for diagnosis and treatment of
XX PT neurological disorders, immunological disorders including autoimmune/
XX PT inflammatory disorders and cell proliferative disorders such as cancer

Matches	192;	Conservative	115;	Mismatches	264;	Indels	172;	Gaps	27;
Qy	504	LTESLTSRGNKARGLQHS	ISVDLSSLSSTLSNTSKPSVCEALPISESSKLL	563					
Db	311	ipeavnvgsacfkafqtppsmcs	insavdmsvts-----dvsgnplntagye	364					
Qy	564	GS-----SEDLSDDSSHLP	PEEPAPLSPOQAFRRRANTLUSHPDCEQPP	612					
Db	365	vsleirervaknsyaavpr	-----fgcfklrantdkvecltqktpsnvlgphierc	418					
Qy	613	---GSPG--VSQR--KL	MRVHSYSTPHERKDFESKANHLGDSGGTTPVK	657					
Db	419	fgvlvapgklvqvkdmlhdm	smgytqp-----ggtgvatesdsnaq	462					
Qy	658	RHSRQOIFLRVANPQKAC	SSRYEDYSELG---ELPPRSPLBPV	700					
Db	463	sswpytiraewkaqekafeq	lniesaktnltvavdivmrirgepvrfvietptigas	522					
Qy	701	---CEDGFGPPPE-----	EKKTT-----SRELRUWQAKILQOILLRMEK	739					
Db	523	emrilmhfmskrptnlrf	lhlkrtesnwkvnsidpseeite--qpghgssllkmgm	580					
Qy	740	EN-----OKLOASENDL	LNLKLLDY-----EETPLCKEYTTWVKMLS	779					
Db	581	nnlsrivrssiasiedcps	-----dyssdgdplsgtgevskdcsgdtldewdpilr	635					
Qy	780	TPGRSKTKFOMEK----	MHSAVGCGVPRHRHGETWKFLEQFHLKHQFPSPKQPKDVPYK	835					
Db	636	-----ewdsekrpknl	aplrvlgvpealrekvwklla-----nvegrmemd-kyk	680					
Qy	836	ELLKOLTSQQOHAILED	IGRTFTHPYFSQAQAGQLSLYNILKAYSLDDQEVGYCQGLSF	895					
Db	681	ilicketkctetvigrd	ihrtfpahkcfeigsgsgdalfkvskayavhdsevgycqglaf	740					
Qy	896	VAGILLHMSSEEAFAKML	KPLMPDNGLRKQYRPMIILQIMYOLSRLLHDYHRDLNHL	955					
Db	741	laaslllhmpdeafval	mydygrldrykagfevrlrylglerlikdqlpkhehf	800					
Qy	956	EEHEIGPSLVAAPFLM	FAFASQPLGFVARVDFMIFLQTEVIFKVALSLGSHKPLIQ	1015					
Db	801	tacgiehmyasqwlly	tarfplcfvfhvlgvllidgipvlqavatl--sicesdlr	859					
Qy	1016	HENLETIVDFIKSTLPN	--LGLVOMETINQVFMEDIAQLQAYEVEHYLQELIDSSP	1073					
Db	860	qldfegilkyfrvtl	pkccrssqarkvmkqacerk--kklkyeef-----	906					
Qy	1074	LSDNORMDKLEKTN	SLRKQNLDLLEQLQVANGRI---QSLEATIEKLLSSESKLQAM	1129					
Db	907	llkkqhkerlek-----	eaqlyenrfgeermkmaeidalnqltsakera	952					
Qy	1130	LTLEERSALLQTV	EEELRRSAE 1152						
Db	953	vekekhtgtlqeyk	qlirgeq 975						
RESULT	15								
AA	AAyl1605								
ID	AAyl1605	standard; Protein;	94 AA.						
XX	AAyl1605;								
AC	AAyl1605;								
XX									
DT	16-JUN-1999	(first entry)							
XX									
DE	Human 5' EST	secreted protein	SEQ ID NO:257.						
XX									
KW	Human; secreted protein; EST;	expressed sequence tag; diagnosis; forensic; gene therapy; chromosome mapping; signal peptide;							
KW	upstream regulatory sequence; cytokine activity; cell proliferation;								
KW	differentiation; haematopoiesis regulation; tissue growth regulation;								
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;								
KW	thrombolytic; anti-inflammatory; tumour inhibition.								
XX									
OS	Homio sapiens.								

XX	WO9906439-A2.	
XX	11-FEB-1999.	
XX	31-JUL-1998; 98WO-TB01233.	
XX	01-AUG-1997; 97US-090468.	
XX	(GEST) GENSET.	
XX	Duclert A, Dumas Milne Edwards J, Lacroix B;	
XX	WPI; 1999-153700/13.	
XX	N-PSDB; AAX40323.	
XX	New nucleic acids encoding human secreted proteins - obtained from	
PT	cDNA libraries derived from liver, lung, large intestine, colon,	
PT	thyroid and pancreas tissue	
XX	Claim 27; Page 351-352; 398pp; English.	
XX	AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for	
CC	human secreted proteins, and encode the proteins given in AAY11533 to	
CC	AAY11679, respectively. The proteins given represent the signal peptide	
CC	and an N-terminal fragment of a secreted protein. The nucleic acid	
CC	sequences can be used for producing secreted human gene products. They	
CC	can also be used to develop products for diagnosis and therapy. The	
CC	proteins obtained may have cytokine activity, cell	
CC	proliferation/differentiation activity, haematopoiesis regulating	
CC	activity, tissue growth regulating activity, reproductive hormone	
CC	regulating activity, chemotactic/ chemokinetic activity, haemostatic and	
CC	thrombolytic activity, receptor/ ligand activity, anti-inflammatory	
CC	activity, tumour inhibition activity or other activities. The products	
CC	can be used in forensic, gene therapy and chromosome mapping procedures.	
CC	The sequences can also be used for obtaining corresponding promoter	
CC	sequences. The nucleic acids encoding the signal peptide can be used for	
CC	directing extracellular secretion of a polypeptide or the insertion of a	
CC	polypeptide into a membrane, or importing a polypeptide into a cell.	
XX	Sequence 94 AA;	

Query Match 7.9%; Score 481; DB 20; Length 94;
Best Local Similarity 97.9%; Pred. No. 7.3e-31;
Matches 92: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	MEPTFTARKHLLPNEVSVDFGLQVLVSLVPHVSTTTHPMLPWVAEVRRLSRQSTRKEPV	60
Db	1	mep1ftarkhllpnevsvdfglqlvslpvhvsttthpmlpwvavvrlsrqstrkepv	60
Qy	61	TKQVRLCVSPSGLRCEPEGRSQQWDPLIYSIF	94
Db	61	txarlcvspsglrcepegrsqwdplivsisf	94

Search completed: August 28, 2002, 16:23:49
Job time: 82 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:22:27 ; Search time 16.95 Seconds
(without alignments)
1683.132 Million cell updates/sec

Title: US-09-762-311-5
Perfect score: 6055
Sequence: 1 MEPTTFARKHLLPNEVSVD.....RSAPSDREPECTQPEPTGD 1168

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents:AA.*

- 1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4976	82.2	1141	US-08-363-300-2	Sequence 2, Appl1
2	232.5	3.8	3248	US-08-353-700-1	Sequence 1, Appl1
3	232.5	3.8	3248	PCT-US95-16216-1	Sequence 1, Appl1
4	211	3.5	2482	US-08-328-254-6	Sequence 6, Appl1
5	197	3.3	376	US-08-253-155A-33	Sequence 33, Appl1
6	196.5	3.2	1093	PCT-US93-03077-1	Sequence 1, Appl1
7	190.5	3.1	885	US-08-533-306A-4	Sequence 4, Appl1
8	190.5	3.1	885	US-08-742-923A-4	Sequence 4, Appl1
9	180	3.0	1886	US-08-938-105-3	Sequence 3, Appl1
10	175	2.9	1939	US-09-310-187A-1	Sequence 1, Appl1
11	169	2.8	816	US-08-533-306A-6	Sequence 6, Appl1
12	169	2.8	816	US-08-742-923A-6	Sequence 6, Appl1
13	160.5	2.7	1312	US-08-687-080-51	Sequence 51, Appl1
14	159.5	2.6	1312	US-08-592-126-148	Sequence 148, Appl1
15	156.5	2.6	1001	US-09-060-410-2	Sequence 2, Appl1
16	153	2.5	2101	US-08-466-390-4	Sequence 4, Appl1
17	153	2.5	2101	US-08-470-950-4	Sequence 4, Appl1
18	153	2.5	2101	US-08-467-781-4	Sequence 4, Appl1
19	153	2.5	2101	US-08-195-487-4	Sequence 4, Appl1
20	153	2.5	2101	US-08-483-924-4	Sequence 4, Appl1
21	153	2.5	2101	US-09-452-294-1	Sequence 1, Appl1
22	153	2.5	2101	PCT-US93-06160-4	Sequence 4, Appl1
23	151.5	2.5	835	US-08-368-751-4	Sequence 4, Appl1
24	151.5	2.5	1388	US-09-572-191-2	Sequence 2, Appl1
25	150	2.5	1164	US-09-457-708-2	Sequence 2, Appl1
26	148.5	2.5	1354	US-08-685-871-2	Sequence 2, Appl1
27	142.5	2.4	1104	US-08-923-992A-4	Sequence 4, Appl1

28	141.5	2.3	2548	4	US-09-172-422-1	Sequence 1, Appl1
29	138	2.3	988	2	US-08-286-819A-19	Sequence 19, Appl1
30	138	2.3	988	3	US-08-980-357-19	Sequence 19, Appl1
31	136.5	2.3	1122	2	US-08-619-198-3	Sequence 3, Appl1
32	136.5	2.3	1389	2	US-08-619-198-5	Sequence 5, Appl1
33	135	2.2	557	4	US-08-979-608A-5	Sequence 5, Appl1
34	134	2.2	1618	1	US-07-853-913-4	Sequence 4, Appl1
35	132.5	2.2	710	4	US-09-079-812E-2	Sequence 2, Appl1
36	132	2.2	2485	4	US-09-290-640-46	Sequence 46, Appl1
37	131	2.2	976	4	US-09-104-324B-4	Sequence 4, Appl1
38	130.5	2.2	1375	4	US-09-722-139-2	Sequence 2, Appl1
39	130.5	2.2	1863	2	US-08-603-753D-2	Sequence 2, Appl1
40	130.5	2.2	1863	4	US-09-099-753-2	Sequence 2, Appl1
41	130.5	2.2	1863	4	US-08-986-106-2	Sequence 2, Appl1
42	130.5	2.2	1863	4	US-09-007-678B-49	Sequence 49, Appl1
43	130.5	2.2	1898	1	US-08-056-200-94	Sequence 94, Appl1
44	130.5	2.2	1898	2	US-08-800-644-94	Sequence 94, Appl1
45	130.5	2.2	1976	3	US-09-024-020B-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-08-363-300-2
; Sequence 2, Application US/08363300
; Patent No. 5700927
; GENERAL INFORMATION:
; APPLICANT: Zon, Leonard and Richardson, Paul
; TITLE OF INVENTION: Tbcd Gene and Uses Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,300
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04590/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-363-300-2

Query Match	82.2%	Score	4976	DB 1	Length	1141
Best Local Similarity	87.9%	Pred. No.	0			
Matches	970	Conservative	47	Mismatches	74	Indels 12; Gaps 5
Qy	37	MPMLPWVAEVRRLSRQSTRKEPVTQVRLCVSPSGRLCEPEGRSQOWDPLIYSSIFEC	96			
Db	1	MPMLPWVAEVRRLSGQCSKKRPVTQVRLUWVSPGLRCEPDLEKSPQWDFLTCSSIFEC	60			
Qy	97	KPQVRHKLHNSHDPVGFACLIKEDAVHRQISICVYFKADQTKVPELISSIRQAGKIARQ	156			


```

Db 61 KPORVHKLHNSHDPVSFACLIKEDAAHROSICVYFKADDOTKVPETISSIRQAGKIARQ 120
QY 157 BELHCPSEFDDTSKFEVLFCGRVVAHKKAPPALIDECIEKFNHVSGRGSESRPNP 216
Db 121 EELRCPSEFDDTFAKFEVLFCGRVVAHKKAPPALIDECIEKFNHVSGRGRTD----- 174
QY 217 PHAAPTGS-OEPVRRPMKRSFSPGLRSLAFRAKELDGGRLSSGFTSSPEESDIENHLIS 275
Db 175 -WEAPTGPSAPGPRPKRSFSPGLRSLAFRAKELDGGRLSS-TFSSP-DNDIENHLIG 231
QY 276 GHNVOPDIEENRTMLFTTGSEVYLISPDTKKIALEKNFKIEISPCSGOIRHVDHFGFI 335
Db 232 GHNVOPDMEENRTMLFTTGPSEVYLISPDTKKIALEKNFKIEISPCSGOIRHVDHFGFI 291
QY 336 CRESSGG--GGFHVCYVFOCTNEALVDEIMMTLKQAFVAAVOOTAKAPALCEGCPLO 393
Db 292 CRCSSGGSGGFHVCYVFOCTNEALVDEIMMTLKQAFVAAVOOTAKAPALCEGCPLO 351
QY 394 SLHKLCERIEGMSSSKTKLEQLKHLTLTNQEQATIFEVQKLRPRNEORENELIISFLR 453
Db 352 GLHKLCERIEGMSSSKTKLEQLKHLTLTNQEQATIFEVQKLRPRNEORENELIISFLR 411
QY 454 CLYEERKQEHIIHGMKQTSQMAENIGSELPPSARFRFLDMLKNKAKRSLTLESLSILS 513
Db 412 CLYEERKQESHHTGAPKQTLQVAENIGSPLPPSASRFRDLSLKNRAKRSLSLESLSILS 471
QY 514 RGNKARGLOEHSISVDLSSLSSTLNTSKPSVCEKEALPISESKLGGSSDLSSDS 573
Db 472 RGNKARGLOHSASVDLSSLSSTLNTSKELSMGKEAPVPSSETFKLLGSSDDLSSDS 531
QY 574 ESHLPPEPAPLSPQAFRRRANTLSHFFIECQPPPPARGSPGVSKLMRYHSVSTETP 633
Db 532 EGHIAESALLSPQAFRRRANTLSHFFIECQPPPPARGSPGVSKLMRYHSVSTETP 591
QY 634 HERKDFESKANHLGDSGGTPVKTRRHSWROQIFLRVATPQKACDSSSRVEDYSELGELPP 693
Db 592 HERKDFESKANHLGDSGGTPVKTRRHSWROQIFLRVATPQKACDSSSRVEDYSELGELPP 651
QY 694 RSPLEPCEDGPGPPPEEKKRTSRELRLWOKALIQIILLRWKENOKLOASENDLLN 753
Db 652 RSPLEPCEDGPGGYKKRGRHASFEGCKRPSRCLVRWKENOKLOASENDLLN 711
QY 754 KRLKLDYEETPCLEKVTWWEKMLSTPGRSKIKFDMKMSHSAVGQGVPRHRHGEIWKFL 813
Db 712 KRLKLDYEETPCLEKVTWWEKMLSTPGRSKIKFDMKMSHSAVGQGVPRHRHGEIWKFL 771
QY 814 AEOFHLKHQPPSKQOPKQDVPYKELLKQLTSQQAAILDILGRTTPTTHPYFSAQLGAGQLSL 873
Db 772 AEOFHLKHPPSKQOPKQDVPYKELLKQLTSQQAAILDILGRTTPTTHPYFSAQLGAGQLSL 831
QY 874 YNLIKAYSLLDQEVGYCOGLSFVAGILLHMSBEEAFKMLKFLMDGLRKQYRPMIIL 933
Db 832 YNLIKAYSLLDQEVGYCOGLSFVAGILLHMSBEEAFKMLKFLMDGLRKQYRPMIIL 891
QY 934 QIQMYQLSRLLDHYHRLDYNHLEBEHETGSLYAAPWFLTMFASQFPLGFGVARVDFMIFLQ 993
Db 892 QIQMYQLSRLLDHYHRLDYNHLEBEHETGPTTAAAPWFLTMFASQFPLGFGVARVDFMIFLQ 951
QY 994 GTEVIFKVALSLGSHKPLIQLHENLETIVDFIKSTLPNLGLVQMEKTIINQVEMDIQAK 1053
Db 952 GSEVIFKVALSLGSHKPLIQLHENLETIVDFIKNTLPNLGLVQMEKTIISQVEMDIQAK 1011
QY 1054 LQAEVEHYHVLQELIDSSPLSDNORMDKLEKTNSSLRKONLDLLEQLQVANGRIQSLA 1113
Db 1012 LQAEVEHYHVVQELIESPLSDNORMDKLEKTNSSLRKONLDLLEQLQVANARIQSLA 1071
QY 1114 TIEKLLSSESKLQAMLTLELER 1136
Db 1072 TVEKLLTSESKLQALTLLEVER 1094

```

RESULT 2
US-08-353-700-1

```

; Sequence 1, Application US/08353700
; Patent No. 559919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TRANSLANTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

```

Query Match 3.8%; Score 232.5; DB 1; Length 3248;
Best Local Similarity 19.3%; Pred. No. 2e-09;
Matches 240; Conservative 191; Mismatches 440; Indels 373; Gaps 55;

```

QY 74 RCEPEPGRSQWDPDIYSSIFECK-PQRVHKLHNSHDP-S-YFACLIKEDAVHRSICVY 131
Db 120 RCKSELSRQQAASADVSLNCPNTQ---KIFTPLTPSYSGSKYEDLKEK-----YN 172
QY 132 FKADDTKVPETISSIRQAGKIARQBELHCPSEFDDTSKFEVLFCGRVTVAAHKAPPA 191
Db 173 KEVERKRLAEAYKALQ-----AKK-----ASQTLQA 200
QY 192 LIDCEKFNHVSGRGSESPRPNPHAAPTGQEPVRRPMKRSFSPGLRSLAFRAKEL 251
Db 201 TMHRIARHQASVSVSWQEQETPSHLSSNSQRTPIRRDFSASYFSGELEVPSTRTLQ 260
QY 252 DGLL-RSSGFFSS-----FEESDIENHL-----ISGHNVOPDIEENRTMLF 293
Db 261 IGRDANSFPFGNSSPHLLDQLKAQNOELRNKINELELRLOGHEKMKQVANKFBLQL 320
QY 294 TIQSEVYLISPD-----TKKIALEKNFKIEISPCSGOIRHVDHF 332
Db 321 QLEKAKVELIEKEKLVNKRDELVRTTTAAQDOASTKYTALEQLKLT-----E 369
QY 333 GFICRESSGGGFHVCYVFOCTNEALVDE-----IMMTLKQAFVAAVOQ 378
Db 370 DLSCQRQNAESA-----RCSLEQKIKEKEFEQBELSRQORSFOTLQ-----ECIQM 417

```

```
QY 379 TAKAPAOQCEGCPQSLHLKCE--RIGMNSSKTKLE--LQKHLTTLTNOEQATIFEVOK 435
Db 418 KARLTOELQAA---KNMHNVLQAEKLTQSVKQOOLENNEEFKQKLCRAEQAFQASQIKE 474
QY 436 LRPRNEQRENELIISFLRCLYEKQKEHIHI--GEMKOTSOM--AAENIGSELPPSATRR 492
Db 475 NEURRMEEMKKNLKLKSHSEQKAREVCHLEAELNKOCLNQSQNFQAEEMKAKNTS-Q 533
QY 493 LDMKL-----NKAKRSLT--ESLESILSRGNKARGL-----QHSISVDLSSLSSTL 538
Db 534 ETMLRDLQEKINQENSLTLEKLVADLEKQKDCSQDLLKKREHHI-----EQLNDKL 588
QY 539 SNTSKPSVCEKALPISSEFKLLGSSDLSSDSHLPPEAPLSPQOAFRRRANTLS 598
Db 589 SKTEKES-----KALLSLELKKKEEELKEETLFS-----620
QY 599 HFPIEQEPPQPARPGVSGQRKLMRYHSVTETPERKDFESKANHLGDSGGTVPKTRR 658
Db 621 -----CWKS-----ENEKLL-----TQWSEKENLQSKINHL-----ETCLKTQ 655
QY 659 ----HSW-----RQIFLRVATPQKACSSSYEDYSELGELPPRSPLEPVCEGDP 705
Db 656 IKSHEYNERVRLTMDRENLSVEIRNLHNVLDKSVSEVETQKLAYMELQKAE-----708
QY 706 FGPPPEKKRTSRELWQKATLOQILLRMEKQKQLOASENDLNN-----RLKLDY 760
Db 709 FSDQKHQK-----EIEENMLKTSQLTQGVDEHKLQLLSNEIMDKRQYQDLHAE 760
QY 761 EETTPCLKE-----VTVWEKMLSTPGRSKIKFDMK--KMH-----SAVG--QGVPRHRGEI 809
Db 761 ESLRDLKSKDASLVN-----EDQRSLLAFDQPMHHSFANIIGEGGSMPSERSEC 814
QY 810 WKFLAEQ-----FHLKHOPPSQKQKQVVPYKELLKOLTSQQHAILIDLG 853
Db 815 -RLEADQSPKNSAILQNRVDSLEFSLEQ---KOMNSDL--QKQCEELVQIKGEIENLM 868
QY 854 RTPTTHPYSAQLG-----AQQLS--LYNILKAYSILQDEVYCVQGLSF 895
Db 869 KASQMHOSFVAETSQRISKIQEDTSAHQNVVAETLSALENKEKELEQLLNDKVETEQ----924
QY 896 VAGILLHNSHSE---EBAFKMLKFLMFDGMLGRKQYRQPMIQLQIMYQLSRLLDHYRDL 952
Db 925 -AIEQELKSNHLEDSLEKLOLSETLSLEKKEMSIISL-----964
QY 953 NHLEHEIGPSLYAAPWFLTMFASQPLPGFVARVDFMIFLQGTVEVIFKVALSLLGSHKPL 1012
Db 965 NKREIBEL-----TQENGTLKEINASLNQEKMNL 993
QY 1013 ILQHENLETIVDFIKSTLPNLG-----LYOMEKTIQVPEMDIAKOLQAYEYHV 1063
Db 994 IQKSESFANYIDREKSISELSQYKQEKILLIQRCEETGNAYE--DLQSKYKAAQEKNSK 1052
QY 1064 LOBELIDSSPLSDNQMDKLEKTNSSLRKQNLQDLLOLQVANGRIOSLE---ATIEKLIS 1120
Db 1053 LECLLNCTSLCEN--RKNLEQLKEAFKAEHQEFLKLAFAERNQNLMLLETVOQALR 1111
QY 1121 SE-----SK-----LKOAMLTLELERSALLQTVBELRRRSAB 1152
Db 1112 SEMTDQNNKSEAGGLKQEBIMTLKBEQKNQKQEVNDLLQENQ 1155

RESULT 3
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
```

```
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1
```

Query Match 3.8%; Score 232.5; DB 5; Length 3248;

Best Local Similarity 19.3%; Pred. No. 2e-09;

Matches 240; Conservative 191; Mismatches 440; Indels 373; Gaps 55;

```
QY 74 RCPEPRGSRQMDPLYISSIFECK-PQRVHKLHNSHDPG-YFACLIKEDAVHRSQICV 131
Db 120 RCKSELSERQQAQASADVSLNCPCTPQ---KIFTPLTPSQYSSGSKYEDLKEK----YN 172
QY 132 FKADDTQKYPEIISIRQAGKIARQEBELHCPSEFDDTFSKKFVFLFCGRVTVAHKKAPPA 191
Db 173 KEVEERKRLAEVKAQ-----AKK-----ASQILPQA 200
QY 192 LIDECIEKFNHVSGRGSESPRNPAPHAAPTGSQEPVRPMRKSFSQPLRSIAFARKEQ 251
Db 201 TNNHRDIARHQASSVFSWQOEKTPSHLSSNSQRTPIRRDFSASYSFSEGELEVPSSRTLQ 260
QY 252 DGGL-RSSGFFSS-----FEESDIENHL-----ISGHNIVQPTDIBENRTMLF 293
Db 261 IGRDRANSFFGNSSPHLLDQLKAQNLKRNKINELELRLQHEKEMKQGVNKFQELQL 320
QY 294 TIGQSEVYLISPD-----TKKIALEKNFKEISFCSGIRHVDHF 332
Db 321 QLEKAVELIEKEKVLNCKRDELVRVTAQYDQASTKYTALEQKLKLT-----E 369
QY 333 GFTCRESSGGGFHFVYVFCQTNALVDE-----IMWTLKQATVAAVQ 378
Db 370 DLSCQRQNAESA-----RCSLEQKIKEKEFOBELSRQORSFOTLDQ-----ECIQM 417
QY 379 TAKAPAOQCEGCPQSLHLKCE--RIGMNSSKTKLE--LQKHLTTLTNOEQATIFEVOK 435
Db 418 KARLTOELQAA---KNMHNVLQAEKLTQSVKQOOLENNEEFKQKLCRAEQAFQASQIKE 474
QY 436 LRPRNEQRENELIISFLRCLYEKQKEHIHI--GEMKOTSOM--AAENIGSELPPSATRR 492
Db 475 NEURRMEEMKKNLKLKSHSEQKAREVCHLEAELNKOCLNQSQNFQAEEMKAKNTS-Q 533
QY 493 LDMKL-----NKAKRSLT--ESLESILSRGNKARGL-----QHSISVDLSSLSSTL 538
Db 534 ETMLRDLQEKINQENSLTLEKLVADLEKQKDCSQDLLKKREHHI-----EQLNDKL 588
```



```
Qy 579 E-----BPAPLSPOAFRRRANTLSHPFIECOEPPQARGSPG 616
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1183 EVOLMTKIEACIELEKIVGELKKNDSLEK-----LEYFSCDHQELLQRTVETSEG 1233
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 617 VSQKRLMYHVSVTETPHERKDFSKANHLGDSGGTPVKTRRHSWROQIF-----666
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1234 LNSDLEM-----HADK---SSREDIGDN---VAKVNDMSKFRFLDVENELSRIR 1276
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 667 -----LVATPQKAC---DSSRYEDYSELGELPPRSPLEPVCEGDPFG 707
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1277 SEKASIEHEALYLEADLEVVOYETKLCLEKDNKQKVI-----VCLLEELS 1322
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 708 PPPBEKRTSRELRLQKAILQIILLRMEKENQKQASENDLN-----K 754
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1323 VVTSERNQLRGELDTMSKTTALDQLEKMEKTKQELSHQSECLHCIOVAEAEVKEKTE 1382
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 755 RLKLDYEITPCLEKVTWVKMLSTPGRSKIKFDMKMHSAVGQVPRHRHGEIWKFLA 814
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1383 LQTLSSDVSELLKDKTHLQSKLAS-----LEKDSQALS-----L 1417
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 815 EGFHLKHQFPKQPKVPYKE---LLKQLTSQOH-----AILDLGRTPPTHY 861
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1418 TKCELENOIAQINKEKELVKESELSQARLSSESYKLVNSKALEAALVKEGE-----1470
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 862 FSAQIGAGQLSLYNLIKAYSLL-----883
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1471 FALRLSSQEEVHQLRGIEKLRVRIEADKQLHIAEKLERERENDSLKDKVENLERE 1530
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 884 -----DOEVGYCOGLSPVAGILLHMSEBEAFKMLKFLMFD-----GLRQRYRPD 929
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1531 LQMSBENQELVILDAENSKAEVETLKTQIEEMARSLKVFELDLVTLRSEKENLTKQIQEK 1590
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 930 MIIQIQMYQLSRLHDYRDLYNHLBEHE--IGPSLYAAPWFLTMFASQFPLGFVARVF 987
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1591 -----OGQISELDKLLSSP-KSLLEEKQAEIQIEESKTAVEMLQNLKELNEAVALCG 1645
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 988 DMIFLQGTPE-----VIFKVALSLGSHKP---LILQHENLETIVDFIKST 1029
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1646 DOEIMKATEQSDLPPIEHEQHLRNSIEKLRARLEADEKKQCLVQQLKESEHHDLKGR 1705
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1030 LPNL-----GLVQMEKNTINQVEM-----DIAKOLQAYEYEVHYL---QEELI 1069
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1706 VENLERELEIARTNOEHAALAEANSKGEVETLKAKIEGTQSLRGLELDVVTIRSEKENL 1765
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1070 DSSPLSDNORDMKLEKTNS-----LRKQNLDLLEQLQV-----ANGRIOSL 1111
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1766 TNEIQKEQERISELTIINSFENILQKEQEKVQMKESSTAMEMLQTLKELNERVAAL 1825
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1112 EATIEKLLSSSKLQKQAMLTLELERSALLQTVIELR 1147
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1826 HNDQEAQKAEQNLSSQVECLEKEKQALLOGLDEAK 1861
| : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 5

US-08-253-155A-33

Sequence 33, Application US/08253155A

Patent No. 5691147

GENERAL INFORMATION:

APPLICANT: Gyuris, Jenő

APPLICANT: Draetta, Giulio

TITLE OF INVENTION: CDK4 Binding Proteins

NUMBER OF SEQUENCES: 95

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 State Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/253,155A
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-253-155A-33
```

Query Match 3.3%; Score 197; DB 1; Length 376;

Best Local Similarity 23.9%; Pred. No. 5.5e-08;

Matches 66; Conservative 41; Mismatches 99; Indels 70; Gaps 6;

```
Qy 771 TTVVEKMLSTPGRSKIKFDMKMHSAVGQVPRHRHGEIWKFLAEQFHLKHQFPKQPK 830
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 TSKWMEML---GEWETKYHSSKLDIVYKIGIPMNIRGPVSVLLNIQIKLNPGRYO-- 128
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 831 DVPYKELLKQLTSQOHAILDLGRTPFPHYFSAQIGAGQLSLYNLIKAYSILLDOEVGYC 890
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 --INKERCKRSESHHIDLDVTRTLRNHVFRRDYGAKQRELFVILLAYSEYNPEVGYC 186
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 891 QGLSFVAGILLHMSEBEAFKMLKFLMFD-----MGLRQYRPDMILQ 934
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 RDLSHITALFLYLPEEDAFWALVQLLASERHSRHPGFHSPNGTVOGLQDQ-----237
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 935 IQMQLSLLHDYRDLYNHLHEEIGPSLYAAPW-----FLTMFASQF 978
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 -----OEHVVPKSPQKTMHQDKGLGCGQCASLGCLLRNLIDGI 276
| : : : : : : : : : : : : : : : : : : : : : : : :
Qy 979 PLGFVARVDMIFLQGTVEVFKV-ALSLLGSHKPLI 1013
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 SLGLTLRLADVLYVEGEQVLMPTISIALKVQOKRLM 312
| : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 6

PCT-US93-03077-1

Sequence 1, Application PC/TUS9303077

GENERAL INFORMATION:

APPLICANT: Board of Regents, The University of Texas System

APPLICANT: Gaynor, Richard B.

APPLICANT: Wu, Foon Kin

TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR

TITLE OF INVENTION: REGULATING GENE EXPRESSION

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/03077

FILING DATE: 19930331

CLASSIFICATION:

Db 507 DAKLREV-----NQALGQFERDLQARDEQNEEKKRQLQQLHETEL-----EDERN 557
Qy 862 FSAQLGAGOLSYNLIKAYSLDDQVGYCGLSFVAGILLHMSBEEAPKMLKFLMFMG 921
Db 558 ERALAAAARKKLEGLDKLEL-----QADSAIKG-----REEAIKQ-----593
Qy 922 LRKQYRPMIILQIQMYQLSRLLDYHR---DLYNHLEBE-IGPSLYAAPWFLTMFASQ 977
Db 594 LKK-----LOAQMKDFORELEADARSDEIFATAKENKAKSLEA-----DIMQLQ 640
Qy 978 FPLGFVARVDFMIFLOGTEVIFKVALSLGSHKPLIQHE--NLETIVDFIKSTLPNLGL 1035
Db 641 EDLAAARARKQADLEKELEBELASSLSGRN---ALQDEKRRLEA-----RI 685
Qy 1036 VQMEKTINQVFMINDIAKQLQAYEYHVLEQELIDSSPLSDNQMDKLEKTNSSLRKQML 1095
Db 686 AQLEEELE---EQQMEAMSORVRKATQAOQLSNELATERSTAQKNESARQQLERQNK 742
Qy 1096 DLLEQLQVANGRIQS-----LEATI-----EKLSSSESKLQOAM 1129
Db 743 ELRSKLEHEGAVKSKFTIAALEAKIAQLEEQVEQAREKQATKSLKQDKKLKEIL 802
Qy 1130 LTLERSALLQVTEELRRRSAPSDREPECTQPE 1164
Db 803 LOVEDERKMAEQYKEQAKGNARVKLQKQLEAE 837

RESULT 9

US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Leinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,105
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crook, Wagnell M.
; REGISTRATION NUMBER: 31,071
; REFERENCE/DOCKET NUMBER: 13595-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1886 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-938-105-3

Query Match 3-08; Score 180; DB 4; Length 1886;
Best Local Similarity 18.7%; Pred. No. 1.7e-05;
Matches 161; Conservative 135; Mismatches 320; Indels 246; Gaps 33;

Qy 422 TNOEQATIFEVOKLR---PRNEQRENEL---IISFLCLRYBEKKEKHIHIGEMKQTSOM 475
Db 794 TEKEMANKKEEGRVKDALEKSEARKKELEKMWSL-----QEKNDLQLOV-QAQODNLA 848
Qy 476 AAENIGSELPPSATRFLD-MLKNKAK-----RSITSELESILSRG-----NKARGLOEH- 524
Db 849 DAE-----RCQLIKNKIQLEAKVKEMTERLEDEEMNAELTAKRKRLEDEC 896
Qy 525 -SISVDLSSLSSTLSNTSKPSVCEKEALPISEBSFKLLGSSEDLSSPSESHLPEPAP 583
Db 897 SELKKDID-DLETTAKVEKEKHATENKVNKTEEMAGLDEIIAKLTKEKKAQLEAHQA 955
Qy 584 LSPQOAFRRRANTLSHFPIECQEPQPARSPGVQSKMLRY-----625
Db 956 LODLAQAEEDKVTNLKSKVKLEQQVDDLBS--LEQEKVMDLERAKRKRLEGLKLTQE 1013
Qy 626 -----HSVTETTERKDFESKANHLGDSGTPVKTRRHSWRQOIFURVATPQKACDS 678
Db 1014 STMDLENDKQLQLEELKKKEFD-----ISOQNSKIEDEQALALQKKLKN 1060
Qy 679 SSRYEDYSFELGELPPRSPLEPVCEGPGPPPEKKRTSRELRLWQKAILQOILLRWE 738
Db 1061 QARIELEE--EL-----EARTAKVEK-----LRSDLTRELE 1093
Qy 739 KENOKLQ-----ASENDLLNKLKLDYEBITPCLKEVTTWKEKMLSTPGRSKIKFDMEXM 793
Db 1094 EISERLEEAGGATSVQIENMKKREAEFOKMRDRLEATL-----Q 1133
Qy 794 HSAVGQVPRHRGELWKFLEAQFHLKHQFPKQKQDPVPYKELLKQLTSSQOHA1-----848
Db 1134 HEATAAALKRKHADSVAE--LGEQIDNLRVKQKLEKSEFKLEDDVTSHMEQIIKAKA 1192
Qy 849 -LIDLGRTFPTH-P-YFSAQLGAGOLSLYNILKAYSLDDQVGYCGLSFVAGILLHMS 906
Db 1193 NLEKVSRTLEDOANEYRVKLEEAQRLNDFTTORAKLTE-----NGELARQLEE 1242
Qy 907 EEA-----FKMLKFLMDMG-----LRKOYRPM-- 930
Db 1243 KEALIWOLTRGKLSVTQOMEDLKQLEEBGKAKNALAHALQSAHRHDCDLLREQYEBEEMA 1302
Qy 931 -----IILQIQMYQLSRLLDYHRDLYNILEBEHIGPSLYAAPWFLTMFASQPLGFVAR 985
Db 1303 KAEQRLVRSKANSEVAQWRTKYETDAIQTREELEAKKLAQR-----1345
Qy 986 VEDMIFLOGTEVIFKVALSLGS-----HKPLILHENLETIVDFIKSTLPNLGLVQMEK 1040
Db 1346 -----LODAEAEVAVNAKSSLEKTKHR---LQNEIEDLVQDVERSNAAAAADKKOR 1396
Qy 1041 TINQVFEMDIAKOLQAYEYHVLEQELIDSSPLSDN-----ORMDKLEKTN 1088
Db 1397 NFDKI---LAEWKQKYESQSELESSQKARSLSLEFKLNAYEESLEHLETFKRENK 1452
Qy 1089 SLRKONLDLLEQLQVANGRIQSLIETIEKL-----LSSESKLQAMLTLEERSALL---- 1140
Db 1453 NLQEEISDLTEQLEGEGKNVHELEKIRKQLEVEKLEQSALEEAASLEHEGKILRAOL 1512
Qy 1141 ---QTVEELRRRSARPSDREPE 1159
Db 1513 EFNQIKAEIERKLAE-KDEEME 1533

RESULT 10
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12

;; NUMBER OF SEQ ID NOS: 3
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1
;; LENGTH: 1939
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match 2.9% Score 175; DB 4; Length 1939;
Best Local Similarity 18.1% Pred. No. 4.4e-05;
Matches 150; Conservative 139; Mismatches 358; Indels 184; Gaps 27;

Qy 422 TNQOATIFEEVQKLR---PNEORENELIISFLRCLYEK-----QKEHIHIGEMKQT 472
Db 847 TEKEMATKKEEFGRIKTEKSEARRKELEKMYSLQEKNDLQLOVQAEQDNLEAER 906
Qy 473 SOMAAENIGSELPPSATFRDLMLNKAARSLTESILSRGNKARGLOEH--SISVDL 530
Db 907 CDQLIKN-----KIQLEAKVKEMNERLEDEEEMNAELTAKKRKLEDECSLKKDI 956
Qy 531 DSSLSSTLSNTSKPSVCEKALPISSESKLGGSSDLSDSESHLPPEAPLSPQOAF 590
Db 957 D-DLELTAKVEKKEKATENKVNKLTEEMAGLDELIATKTKKALQEAHQALDDQLVE 1015
Qy 591 RRRANTLSHFPIEQEPQPARGSPGVYSORKLMRY----- 625
Db 1016 EDKVNLSKSKVKLEQQVDDLEGS--LQEKVKVRMDLERAKRKLEGDLKLTQESIMDLEN 1073
Qy 626 HSVTETPHERKDFESKANHLGDSGGTPVKTRRHSWQOIFLRVATPKACDSSSRVEDY 685
Db 1074 DKLQLEKKLKKFED-----INQNSKIEDEQALALOKKLEKNAARTEEL 1120
Qy 686 SE--LGELPPRSPLEPCVEDGPPPPPEKKRTSRRELRLMOKAILQOILLRMEKENQK 743
Db 1121 EEELEAERTARAKVKELRSD-----LSRELEEISE-----RLEEAGG- 1157
Qy 744 LQASNDLLNRLKLDVEEITPCLKEVTTWVKMLSTPGRSKIFDKMEKHSVAGQGVPR 803
Db 1158 -ATSVQIEMNKKRAEFQKMRDLDEATL-----QHEATAAALRK 1196
Qy 804 HHRGEIWKFLAEQFHLKHQFSPKQPKDVPYKELLKOLTSOHA1-----LIDLGRTP 857
Db 1197 KHADVAE-LGEQIDNLRQVKQLEKSEKPEKLELDDVTSNWEQIIRAKANLEKVSRTLE 1255
Qy 858 THP-YFSAQLGAGQSLNLIKAYSLLDQEVGYCOGLSFVAGIILLHMSSEAF----- 910
Db 1256 DQANEYRVKLEEAQRSLNDFTTORAKLQTE-----NGELARQLEEKALISQLTR 1305
Qy 911 KWLAFMFDMLGRKQYRPFDMILIQIOMYOLSRLLHDYH--RDLYNHLFE--HEIGPSIYA 966
Db 1306 GKLSYTOQMEDLKKQLEEEGAKNALAHALQASARHDCOLLREQYEETEAKAELQVLSK 1365
Qy 967 APFWLTMPASQFPLGFVARVDFMIFLQGVTEVIFKVALSLGSHKPL----- 1012
Db 1366 ANSEVAQRWRTYETDAIQTEEL-----EAKKKLAQLQDAEAEAVNAKCSLEKTK 1420
Qy 1013 -ILQHENLETTVDFTKSTPLNGLVQMEKTTNQVFEMDIQALQAYEVEYHVILQELIDS 1071
Db 1421 HRLQNEIEDLVMDVRSNAAAALDKKORNFDK1-----LAEWKQKYEESQSELESSOKEA 1476
Qy 1072 SPLSDN-----QRWDKLEKTNSSIRKQNLDLLEOLQVANGRIQSIQLEATIEKL- 1118
Db 1477 RSLSTELPKLNAYEESLEHLETFKRENKNLQOEISDTEQLGEGGKNVHELEKVRKOLE 1536
Qy 1119 ---LSSEKLLKQAMLTLELSALI-----QTVEEELRRRSABSPSDEPE 1159
Db 1537 VEKLEQLQALBEAEASLEHEBKILRAQLEFNQIKAEIERKLAE-KDEEME 1586

RESULT 11
US-08-533-306A-6
; Sequence 6, Application US/08533306A

;; Patent No. 5837457
;; GENERAL INFORMATION:
;; APPLICANT: Liu, Pu
;; APPLICANT: Collins, Francis S.
;; APPLICANT: Siciliano, Michael J.
;; APPLICANT: Claxton, David
;; TITLE OF INVENTION: Markers for Detection of Chromosome 16
;; TITLE OF INVENTION: Rearrangements
;; NUMBER OF SEQUENCES: 14
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
;; STREET: P.O. Box 828
;; CITY: Bloomfield Hills
;; STATE: MI
;; COUNTRY: USA
;; ZIP: 48303
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/533,306A
;; FILING DATE: September 25, 1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Smith, DeAnn F.
;; REGISTRATION NUMBER: 36683
;; REFERENCE/DOCKET NUMBER: 2115-00869COB
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (810) 641-1600
;; TELEFAX: (810) 641-0270
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 816 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 2.8% Score 169; DB 2; Length 816;
Best Local Similarity 19.9% Pred. No. 3.6e-05;
Matches 170; Conservative 112; Mismatches 321; Indels 250; Gaps 34;

Qy 381 KAPALQCEGCPQLSHLKCERIEGMSKTKLEKQKHLTTLTNEQATIFEEVQKLRPN 440
Db 97 KAPMILANGVCYVIWKGWIDLQRLDGMGCLEFDEERAQOEDALAQQ-----AFEEA-----RR 147
Qy 441 EORENELIISFLRCLYEKKEKEHIGEMKOTSQMAAENIGSELPPSATFRFLDMLKNKA 500
Db 148 RTRE-----FEDRDRSH----- 159
Qy 501 KRSITSELESILSRGNKARGLOEHSISVDL-SSLSTLSNTSKPSVCEKALPISSESS 559
Db 160 REEMENEVESVTGMLNEAG---KAIKLAKDVASILSLOLQDTQE---LLQEBETROKLNVS 213
Qy 560 FKLLGSSDLSDSESHLPPEAPLSPQOAFRRRANTLSHFPIEQEPQPARGSPGVSYQ 619
Db 214 TKLRQLEEBERNS-LQDQLEDE---MEAKQNLERHISTLN---IQLSD-----SK 255
Qy 620 RKLARYHSVSTETPHERKDFESKANHLGDSGGTPV-----KTRRHSWRO----- 663
Db 256 KKLQDFASTVALBEGKKRFKQKEIENLTQOYEEKAAAYDKLEKTKNRLQQLDLDLVDD 315
Qy 664 ---QIFLRVATPKACDS-----SSRYEDYSELGELPPRSPLEPCVEDGPGFPPPEE 712
Db 316 NQRLVSNLEKKQKQKFDQLLAEKKNISSKYAD----- 348
Qy 713 KKRTSRELRLWQKAI-----LQOILLRL--WEKENQKLOASENDLLNKR--LKLDYEEI 763
Db 349 RDRAEAAREKETKALSARALEEALKEELERTNKMKAEMEDLVSSKDDVGKNVHEL 408

```

QY 764 TPCLKEVTTWVKMLSTPGRSKIKFDMKMHSAVGQGVPRHHRGEIWKFLAEQFHLKHQF 823
Db 409 EKSKRALETQMEEM-----KTQLELEDELOASEDAKLREV-----NQALKGQF 454
QY 824 PSQOQKDVYPYKELLKOLTSQQHAILDIDGRTPTTHPYFSAQAGQSLSYNLKAYSLL 883
Db 455 ERDLQARDEQNEEKRRQLQRLQHEYTEL-----EDERNERALAAAKKLEGDLKDEL- 509
QY 884 DQEVGVCQGLSFVAGILLHMSSEEAFAKMLKFLMDFMGLRKQYRPMIILQIQMYQLSRL 943
Db 510 -----QADSAIKG-----REAAIKQ-----LRK-----LQAMKDFQRE 538
QY 944 LHDYHR---DLYNHLBEHE-IGPSLYAAPWFLTMFASQFPLGFVARVDFMIFLQGTVEIF 999
Db 539 LEDARASRDEIFATAKENEKAKSLEA-----DLMLQLEDLAAAEARKQADLEKEELAE 593
QY 1000 KVALSLGSHKPLILQHE---NLETIVDFIKSTLPNLGLVQMEKTIQVFMEDIAKQLQAY 1057
Db 594 ELASSLSGRN---ALQDEKRRLEA-----RIAQLEBELE---EQQNMMEAMSD 635
QY 1058 EYEVHYVLOEELIDSSPLSDNQRMDKLEKTNSSLRKQNLIDLEQLQVANGRIQS----- 1110
Db 636 RVKATQQAQAEQLSNELATERSTAKNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIA 695
QY 1111 -LEATI-----EKLSSSEKIKQAMLTLELERSALLQTVBELRRRSA 1151
Db 696 ALEAKIAQLEEQVEQAREKQAATKSLKQDKKILKEILLQVEDEKRMKAEQYKQAEKAGNA 755
QY 1152 EPSDRPECTOPE 1164
Db 756 RVQLKRLQLEAE 768

```

RESULT 12

```

US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: NO. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

MOLECULE TYPE: protein
US-08-742-923A-6

Query Match 2.8%; Score 169; DB 2; Length 816;
Best Local Similarity 19.9%; Pred. No. 3.6e-05;
Matches 170; Conservative 112; Mismatches 321; Indels 250; Gaps 34;

```

QY 381 KAPAQICEGPCLOSLLKLCERIEGMNSSKTKLEQLKHLTLTNOEQATIFEYQKLRPRN 440
Db 97 KAPMIUNGVCVWIKWIDQLRDMGCLGPEDEARAQOEDALAQQ-----AFEEA-----RR 147
QY 441 EQENELIISFLRCLYEKKOKEHIHIGEMKQTSQMAAENIGSELPSPSATRFRMLMLNKA 500
Db 148 RTRE-----FEDRDRSH-----159
QY 501 KRSLTSLSILSRGNKARGLQHSISVDLD--SSLSTSLNTSKPSVCKEKALPISSESS 559
Db 160 REEMENEVESVTGMLNEAEG---KAIKLAQDVASLSQLODQOE---LLOEETRQKLNVS 213
QY 560 FKLLGSSSEDLSSDSHLPPEAPLSPQQAFFRRANTLSHFPIECOEPPQPPARGSPGVSO 619
Db 214 TKLQRLSEERNS-LQDOLDEE---MEAKQNLERHISTLN---IQLSD-----SK 255
QY 620 RKLRYHVSYSTETPHERKDFESKANHLGDSGGTPV-----KTRRHSWRQ-----663
Db 256 KKLQDASTVEALEEKKRFQKEIENLTQYEEKAAAYDKLEKTKNRLQOELDDLVDLD 315
QY 664 ---OIFLRVATPOKACDS-----SSRYEDYSELGELPPRSPLEPVCEDGGFGPPPEE 712
Db 316 NQQLVSNLEKKQKQKFDQLLAEEKNISSKYAD-----E 348
QY 713 KRTSRELRELWOKAI-----LQOILLR--MEKENOKLOASENDLLNKR--LKLQYEEI 763
Db 349 RDRAEAREKETKALSARALEEAELEKERTNKLKRAEMEDLVSSKDDYGVKNVHEL 408
QY 764 TPCLKEVTTWVKMLSTPGRSKIKFDMKMHSAVGQGVPRHHRGEIWKFLAEQFHLKHQF 823
Db 409 EKSKRALETQMEEM-----KTQLELEDELOASEDAKLREV-----NQALKGQF 454
QY 824 PSQOQKDVYPYKELLKOLTSQQHAILDIDGRTPTTHPYFSAQAGQSLSYNLKAYSLL 883
Db 455 ERDLQARDEQNEEKRRQLQRLQHEYTEL-----EDERNERALAAAKKLEGDLKDEL- 509
QY 884 DQEVGVCQGLSFVAGILLHMSSEEAFAKMLKFLMDFMGLRKQYRPMIILQIQMYQLSRL 943
Db 510 -----QADSAIKG-----REAAIKQ-----LRK-----LQAMKDFQRE 538
QY 944 LHDYHR---DLYNHLBEHE-IGPSLYAAPWFLTMFASQFPLGFVARVDFMIFLQGTVEIF 999
Db 539 LEDARASRDEIFATAKENEKAKSLEA-----DLMLQLEDLAAAEARKQADLEKEELAE 593
QY 1000 KVALSLGSHKPLILQHE---NLETIVDFIKSTLPNLGLVQMEKTIQVFMEDIAKQLQAY 1057
Db 594 ELASSLSGRN---ALQDEKRRLEA-----RIAQLEBELE---EQQNMMEAMSD 635
QY 1058 EYEVHYVLOEELIDSSPLSDNQRMDKLEKTNSSLRKQNLIDLEQLQVANGRIQS----- 1110
Db 636 RVKATQQAQAEQLSNELATERSTAKNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIA 695
QY 1111 -LEATI-----EKLSSSEKIKQAMLTLELERSALLQTVBELRRRSA 1151
Db 696 ALEAKIAQLEEQVEQAREKQAATKSLKQDKKILKEILLQVEDEKRMKAEQYKQAEKAGNA 755
QY 1152 EPSDRPECTOPE 1164
Db 756 RVQLKRLQLEAE 768

```

RESULT 13
US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427

GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.
INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 2.78; Score 160.5; DB 2; Length 1312;
Best Local Similarity 18.88; Pred. No. 0.00037;
Matches 175; Conservative 165; Mismatches 366; Indels 223; Gaps 36;

Qy 286 EENRTMLFTIGSEVYLIS-----PDTKKIALEKNFKETSCSQGIRHVDHGF 334
Db 135 EIDREMISSLGSKAVLNVIFFCHOEDSNWPLSEGKALKQKDFEISATRIKALETLRQ 194
Qy 335 ICRESSGGGPHF---VCYVQCTNEALVDEIMMTLKAFTVAAVQQTAKAPAQCEGP 391
Db 195 V-RTQGGKVKQYQMLKYLKQYKACEIRDQITSKEAQLTSSKEIVKSYENEL---DP 250
Qy 392 LOS-----LHKICERIEGNSKTKLEQLKHLTTITNOEQATIFEVOKLPRN 440
Db 251 LKNRLKETEHLNLSKIMKLDNIEKALDSRKKOME-----KDNSELEKMEKVFQGT 300
Qy 441 EORENELIISFLRCLYEKKQK---HHIGEMKQTSQMAENICSELPSPATRFRLDWLK 497
Db 301 DEQLNDLYHNHQRVREKERKLVPCHELEKLNKESRLNNOE-KSELLVEOGRLOLQADR 359
Qy 498 N----KAKRSTESLESTLRGNKARG-LOEHSIS-----VDLSDSLSTLSNTSK 543
Db 360 HQEIRARDSLIQSATOLELDGFERGPFPSRQTNFKHLVREQGEAKTANOLMN--- 416
Qy 544 EPSVCEKAL---PISESSFLLG-----SSEDLSSDSSEHLPPEAPLSPQOAFRRRA 594
Db 417 --DFAKETLQKQIDEIRDKTGLGRIELKSEILSKKQNELKNVYELQQLGSSDRI 474

Qy 595 NTLSHFPIEQEPPQPARGSPGVQSQRKLMRYHSYSTETPHERKDFESKANHLGDSGGTPV 654
Db 475 LELDOELIKAEIRLSKAEKNSNVELTKM---EVSISLQ--NEKADLDRFLRKLDQE---M 525
Qy 655 KTRRHSWROQIFLRVATPOKACDS-----SSRYED--YSELGELPPRSPLEPCVEDGPF 706
Db 526 EQLNHHHTTTRQOMELTKDKADKDEQIRKIKSRHSDELTSLLGYFPNKKQLEDWL----- 580
Qy 707 GPPPEEKRTSRRELRLWQKAILQOILLRLMEKENQKIQASENDL--LNKRLKLDYEIT 764
Db 581 ----HKSKEINQTRD-----RLAKNLKELASSEQNKNHINNELKRKEQL- 622
Qy 765 PCLKEVTVMKMLSTPGRSKIKFDMEMKHSVAVGGVPRHHRGETWTFLAQFHLKHQFP 824
Db 623 -----SYEDKLFVCGSQDFESDLDRK-----KEETE- 652
Qy 825 SKOOP----KDPYKELIKQLTSQOHAILEDLGRFTPTHYFSAQLGAGOLSLYNILKAY 880
Db 653 SKORAMLAGATAVYSQFITQLTDENQSCCPVCQRFQI---EAEI- 694
Qy 881 SLLDQEVGYCOGLSFVAGILLHNSBEEAFKMLKFLMDFMGLRKQYRPMI----- 931
Db 695 ----QEV-----ISDLSQSKRL-----APDKLKSTESLKKKRRDEMLGLVPMRQSI 739
Qy 932 --ILQIQWYQLSRLLHDYHRD---LYNHLREHEIGPSLYAAPWFLTMFASQFPLGFVARV 986
Db 740 IDLKEKEIPELUNKLQNVNRDIQRLKNDIEQE-----TLTGTIMPEESAKV 787
Qy 987 -----FDMIFLAGTEVIFKVALSLGSHKPLILQHENLETVDFIKSTLPNLGLV 1036
Db 788 CLTDVTIMERFQEMELKQVERKIAQAAKLGIDLDRTVQOVNQEK-----QEKQHKLDTV 842
Qy 1037 QMKTINQVFEMDIKQLOAYEYHVLOEELIDSSPLSDNQMDKLEKTNSSLRKQNL 1096
Db 843 SSKIELNRKLIQDQEQIQHLKSTNELKSEKQIS--TNLORQLEEQEOTVELSTEVQS 900
Qy 1097 LLEOLOVANGRIQSLEATIEKLLSSESKL 1125
Db 901 LYREIKDAKEQVSPLETTLEKFOQEKEEL 929

RESULT 14
US-08-592-126-148
Sequence 148, Application US/08592126
Patent No. 5821091
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
TITLE OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54
US-08-592-126-148

Query Match 2.6%; Score 159.5; DB 2; Length 1312;
Best Local Similarity 18.8%; Pred. No. 0.00045;
Matches 175; Conservative 165; Mismatches 366; Indels 223; Gaps 36;

QY 286 EENRTMLFTIGQSEVYLIS-----PDTKIALKKNFKKIFSCQSGIRHVDHFGP 334
DB 135 EIDREMISSLGVSKAVLNNVIFCHQEDSNWPLSEGLAKQKQFDEIFSATRYIKALETLRQ 194
QY 335 ICHRESSGGGFHP---VCYVFOCTNEALVDEINMTLKQAFVAAVQOTAKAPQALCEGCP 391
DB 195 V-RQTQGVKQVOMELYLKQKACEIRDQITSKEAQLTSSKEIVKSYENEL---DP 250
QY 392 LQS-----LHKLCERIEGMSSSKTKLEQLKHLTLTNOEQATIFEVEYQKLRPN 440
DB 251 LKNLKEIEHNLSKIMKLONEIKALDSRKKQMB-----KDNSELEKMEKVFGT 300
QY 441 EQRENELISFLCLVEEQKE---HIHIGEMKQTSOMAENIGSELPPSATRFLDMLK 497
DB 301 DEQLNDLYNHQRTVREKRLKYDCHRELEKLNKESRLNQE-KSELLVEQGRLOQADR 359
QY 498 N-----KAKRSLTSLSLSIRGNKARG-LQEHIS-----VLDSSLSLTLSNTSK 543
DB 360 HQHIRARDSLQSLATQLELDFGFGPFSEROIKNFHLVRRQGEAKTANQLMN--- 416
QY 544 EPSVCEKEAL---PISESFKLIG-----SSDLSSESHPLEPAPLSPQAFRRRA 594
DB 417 -DFAKETLQKQIDRDKTKGLRIITELKSEILSKQNELKNVYELQLEGSSDRI 474
QY 595 NTLSHPIECQEPQPARSGPVGSRKLMRYHSVSTETPHERKDPESKANHLGDSOGTPV 654
DB 475 LEDQLELIKAEKLSKAEKNSNYETLKM-----EVLISQ--NEKADLORTLRKLDQE---M 525
QY 655 KTRHRSWROQIFURVATPQKACDS-----SSRYED--YSELGELPPSPLEPVCEDGPF 706
DB 526 EQLNHHITTTQEMLTQKADKDEQIRKIKSRHSBELTSLGYPFNKKQLEDWL----- 580
QY 707 GPPPEKKRTSRELRLNQKAILQQILLRMEKENOKLASENDL--LNKRLKLDYEET 764
DB 581 ---HSKSKEINQTRD-----RLAKLNKELASSEQKNKHNNELKRREEQL- 622
QY 765 PCLKEVTVWEKMLSTPGRSKIKFDMKMHSAVGQVPRHHRGEIMKFLAEQFLKXHP 824
DB 623 -----SSYEDKLVDCVGSODFSDDLRL-----KEIEK-----S 652
QY 825 SKQOP-----KDVVPYKELLKQTSQOHAILEDLGRTPPTHPYFSAQLGAGOLSLYILKAY 880
DB 653 SKQRAMLAGATAVYSQFITQIDENQSCPCVQCVFTQ-----EAL- 694
QY 881 SLLDQEVGYCGSLFVAGILLMLHMBEEAFKMLKFLMDFMGLKRRKQYRPMI----- 931
DB 695 ---QEV-----ISDLQSKRL-----APDKLKSTESLKKKRRDEMGLGVPMRQSI 739
QY 932 --ILQIQMYOLSRLLHNDYHD--LYNHLREHEIGFSLYAAPWFLTMFASQFPLGVFARV 986
DB 740 IDLKEKEIPELRNKLQNVNDIQLKNDIEQE-----TLLGTIMPEESAKV 787
QY 987 -----FDMIFLQGVTEVIFKVALSLGSHKPLILQHENLETIVDFIKSTLPNLGLV 1036
DB 788 CLTDVTIMERFQMLKDVVERKIAQQAALQGLIDLRTVQVQNEK-----QEQKHLDIV 842
QY 1037 QMEKTIINQVFEMDIQALQAYEYEHVQLBELIDSSPLSNQRQMDKLEKTNSSLRKQNL 1096

DB 843 SSKIELNRKLIQDQEQBIQIHLKSTTNELKSEKLOIS--TNLQRRQQLSEOTVELSTEVS 900
QY 1097 LLEQLOVANGRIQSLSEATIEKLSSSESKL 1125
DB 901 LYREINKAKEQVSPLETTLEKTFQKEEL 929
RESULT 15
US-09-060-410-2
Sequence 2, Application US/09060410
Patent No. 6165461
GENERAL INFORMATION:
APPLICANT: Cobb, Melanie
APPLICANT: Hutchinson, Michele
APPLICANT: Chen, Zhu
APPLICANT: Berman, Kevin
TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
TITLE OF INVENTION: THEREFOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,410
FILING DATE: 14-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.421
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1001 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-060-410-2

Query Match 2.6%; Score 156.5; DB 4; Length 1001;
Best Local Similarity 18.8%; Pred. No. 0.00052;
Matches 153; Conservative 111; Mismatches 282; Indels 267; Gaps 32;
QY 410 TKLELQKHITLTNQDQATIFEVEQKLRPNREORENELIISFLRCLY-----EE 458
DB 271 TSELKLMFVLRRPETVLIDLIQRTKDAVRELDNLQYRKMKKLLFQEAHNPAYEAQE 330
QY 459 KQKEHTHIGEMKQTSOMAENIGSELPPSATRFLDMLNKKAKRSLETSLSRGKA 518
DB 331 EEEQDHGGGRTGT-----VNSVGS-----NQSIPTM----- 357
QY 519 RGLQEHISVLDLSSLSLTSLNTSKEPSYCEKALPISESSFKLLGSSDLSSESHL- 577
DB 358 -----SISASSQSSVNSLPOASD-----DKSELDMM-----GDHTVMSNSSVIHLK 400
QY 578 PEP-----PAPLSPOQAARRAN-----TLSHFPPECQPPQPARC 613
DB 401 PEBENTQEBDPRTASAPQSPQVSRHKSHYRNREHFATIRTSASLVTRMQOEQDSEL 460
QY 614 SPQVSQRLMR-----YHSVSTETPHER-----KDFESKANHLGDSGGTPVKTRRH 659

Db	461	REQMSGYKMRROHQOQLMTLENKUKAEMDEHRLDKDLETOQRNFAAEMEKLIKHHQA	520
Qy	660	SWRQOIFLRVATPKACDSSRYEDYSELGPPRSPLEPVCEOGPFPPPPKEKKRSTSRE	719
Db	521	SMEKEAKV-MANEKKAFQOHIQAQQKKEUNSPLESQKREYKLRKEOLKEELNENQSPKK	579
Qy	720	LRELWOKAILLOIILLRMKEKN-QKLQASENDLLNKRKLKLOYEITPCLKEVTVTWKML	778
Db	580	EKQEW-----LSKOKENIQHQAEEEAANLLRROR-QYLE-----ECRFRKRM	623
Qy	779	STPGRSKIRFDMKMHSAVCGQVPRHGRGEIWKFLAEQFHLKHOFPPSQKQKDPVYKELL	838
Db	624	-----LGRHN-----LEQDLYREELNKRQTKDLEHAMILL	653
Qy	839	KQITSQO-----HAILIDGRTPPTHYPFSAQLAGQSLYNILKAYSLDDQEVGYCOGL	893
Db	654	ROHESMQELEFRHLNTIQMRCELIRLQHOTEL-TNOLE-YNKRRERELRRKHV-----	705
Qy	894	SFVAGILLHMSEEAFAKMLPMDGLRKQYRDMILIOIQMYQLSRLLHDYHRDLYN	953
Db	706	-----MEVROQPKSLK-SRELOIKKQFO-DTCKIQTRQYKALR-----N	742
Qy	954	HLEEIEGFPSLYAAPFLTMFASQFPFGVARVDFMIFLOGTEVIFKVALSLGLSHKPLI	1013
Db	743	HLLE-----ITPKSEHKAVL	757
Qy	1014	--LOHENLETIVDFIKSTLPNLGLVOMEKNTINQVFMEDIAKOLAQAYEYHVHLOEELIDS	1071
Db	758	KRLKEQTRKLAIAE-----QYDHSINEMLSTQALRDEQAECQVILKMQOL----	805
Qy	1072	SPLSDNORMDKLEKTNSSLRKO-----NDDLLEQLQVANGRIQSLTEATIKLLSSEK	1124
Db	806	-----QOELELLNAYQSKIKWAQAQAQHBRELEQO-----RVSLRRALLEQ-----K	848
Qy	1125	LQAMITLELERSALLQTVTEELRRRSAPPSDRE	1157
Db	849	IEDEMLAQONERTERISLLRQAQIEAFQSE	881

Search completed: August 28, 2002, 16:23:08
Job time: 41 sec